Morphometric analysis of data inherent in examination by magnetic resonance imaging: importance to natural history, prognosis and disease staging of squamous carcinoma of the oral cavity

Paul W. Boland

St. John's College, Oxford

A Thesis submitted in fulfillment for the degree of
Doctor of Philosophy

Nuffield Department of Surgery
Radiology Research Group
University of Oxford
April 23, 2010

Abstract

Magnetic resonance imaging plays an important yet underutilized role in determining the natural history and prognosis of oral carcinoma.

Depth of tumour invasion is an emergent factor in the oral cancer literature. However, problems exist with the definition of cut-points suitable for inclusion in TNM staging criteria. Statistical methodology represents a possible explanation but is underexplored. In this work, a review of the depth of invasion literature is conducted with emphasis on statistical technique. As well, statistical simulation is used to explore the implications of the of the minimum p-value method.

The results demonstrate that the use of continuous variable categorization and multiple testing is widespread, and contributes to cut-point variability and false-positive tests. Depth, as a predictor of OCLNM and survival, must be questioned.

The volume of tumour invasion is a promising prognostic factor that has not been fully investigated in the oral carcinoma literature. In this work, the volume of tumour invasion is measured on MRI and compared to thickness and maximum diameter in its capacity to predict 2-year all-cause, disease-related and disease-free survival, as well as occult cervical lymph node metastasis prediction. As part of a comprehensive approach, morphometric factors are incorporated into multifactor predictive models using regression, artificial neural networks and recursive partitioning.

It is evident that MRI-based volume is superior all other linear measurements for both occult cervical lymph node metastasis and survival prediction. Artificial neural networks wee superior to all other techniques for survival prediction. There is a case for a unified artificial neural networks model for survival prediction that uses volume, midline invasion and N-stage to determine prognosis. This model can be used to determine individualized probabilities of 2-year survival.

The lateral extrinsic muscles of the tongue lie just beneath the surface of the lateral tongue, yet their invasion is a criterion for T4 classification using the TNM staging system. In this work, the Visible

Human Female is used to conduct an anatomic study of the extrinsic muscles of the tongue. Linear measurement is used to quantify the distance from the surface mucosa to the most superficial muscle fibres of the styloglossus and genioglossus. Further, the lateral extrinsic muscles are poorly demonstrated on MRI. An anatomic atlas of the tongue is fused with MRI images of oral carcinoma to demonstrate lateral muscle invasion.

The results demonstrate that the styloglossus and hyoglossus lie very close to the surface of the lateral tongue, in some cases passing within 1 mm of the surface mucosa. These extrinsic muscles are readily invaded by even small tumours of the lateral tongue. Strict application of the TNM T4a criteria leads to unnecessary upstaging as these carcinomas do not warrant the prognosis and aggressive treatment of Stage IV disease. Extrinsic muscle invasion should be removed as a T4a criterion for the oral cavity. A separate category, T4a (oral tongue) specifying invasion of the genioglossus is also recommended.

This work presented in this thesis is an original contribution to the field of oral cavity cancer research and has determined that there is capacity for improvement in current efforts to determine the natural history and prognosis of oral cavity squamous cell carcinoma.

This thesis is the first to examine the role of statistical methodology in oral carcinoma depth of invasion cut-point variability. Further, this work presents an original approach to the prediction of regional metastasis and survival using advanced multivariate modeling techniques. No other work explored MRI-measured volume using the substantial sample size gathered in this thesis. Finally, this work is the first to demonstrate that lateral extrinsic muscle invasion is an unnecessary component of the T4a (oral cavity) classification criteria and should be reconsidered.

To my Mom and Dad
for their unending love and support
in everything I do

To my fiancé Rikki.
I never could have done this
without you.

Table of Contents

List of Figures

List of Tables

Abbreviations and Acronyms

2D	Two-Dimensional
3D	Three-Dimensional
ACE-27	Adult Comorbidity Evaluation 27
ACS	All-Cause Survival
AJCC	American Joint Committee on Cancer
ANN	Artificial Neural Network
AP	Anterior-Posterior
ASR	Age-Standardized Rate
BS	Briar Score
BW	Band Width
CIN	Cervical Intraepithelial Neoplasia
CLNM	Cervical Lymph Node Metastasis
CN	Cranial Nerve
CP	Complexity Parameter
CPU	Computer Processing Unit
CRAN	Comprehensive R Archiving Network
CT	Computed Tomography
DFS	Disease-Free Survival
DICOM	Digital Communications in Medicine
DPA	Data Protection Act of 1998
DRS	Disease-Related Survival
DTI	Diffusion Tensor Imaging
EBV	Epstein-Barr Virus
ECS	Extracapsular Spread
EFS	Event-free survival
EGFR	Epidermal Growth Factor Receptor
ENT	Ear, Nose and Throat
FDG-PET	Fluorodeoxyglucose - Positron Emission Tomography
FEG	Frequency Selection Gradient
FEV1	Forced Expiratory Volume (1 Second)
FID	Free-Induction Decay
FLAIR	Fluid Attenuation Inversion Recovery
FND	Functional Neck Dissection
FOV	Field of View
FSE	Fast Spin-Echo
GBSG	German Breast Cancer Study Group
GG	Gengioglossus
GH	Geniohyoid
GLM	Generalized Linear Model
GRE	Gradient Echo
GUI	Graphical User Interface
HCV	Hepatitis C Virus
HG	Hyoglossus

HIV	Human Immunodeficiency Virus
HPV	Human papilloma Virus
HR	Hazard Ratio
HU	Hounsfield Unit
IARC	International Agency for Research on Cancer
IL	Inferior Longitudinalis
IQR	Interquartile Range
IR	Inversion Recovery
JPEG	Joint Photographic Experts Group
JRH	John Radcliffe Hospital
KFI	Kaplan-Feinstein Index
KM	Kaplan-Meier
LAN	Local Access Network
LCD	Liquid Crystal Display
LVI	Lymphovascular Invasion
MDT	Multidisciplinary Team
MPR	Multiplanar Reprojection
MRI	Magnetic Resonance Imaging
MRND	Modified Radical Neck Dissection
NCDB	National Cancer Database
NEX	Number of Excitations
NHS	National Health Service
NIH	National Institutes of Health
NPV	Negative Predictive Value
OCLNM	Occult Cervical Lymph Node Metastasis
OCSCC	Oral Cavity Squamous Cell Carcinoma
OIN	Oral Intraepithelial Neoplasia
OMRI	Oxford Magnetic Resonance Imaging
OPG	Orthopantomogram
OR	Odds Ratio
PACS	Picture Archiving and Communications Service
PAS	Patient Archiving Service
Path	Pathology
PE	Processing Element
PEG	Phase Encoding Gradient
PET	Positron Emission Tomography
PET-CT	Positron Emission Tomography - Computed Tomography
PG	Palatoglossus
Pixel	Picture Element
PNI	Perineural Invasion
PPV	Positive Predictive Value
RAM	Random Access Memory
RF	Radiofrequency
RMT	Retromolar Trigone
RND	Radical Neck Dissection

ROC	Receiver Operating Characteristic
ROI	Region of Interest
SCC	Squamous Cell Carcinoma
SE	Spin-Echo
SEER	Surveillance, Epidemiology and End Results
SG	Styloglossus
SJG	Stephen J Golding
SL	Superior Longitudinalis
SND	Selective Neck Dissection
SNR	Signal-to-Noise Ratio
SPECT	Single Positron Emission Computed Tomography
SPF	S-Phase Fraction
SPSS	Statistics Package for Social Scientists
SSG	Slice Selection Gradient
STIR	Short Tau Inversion Recovery
T	Transversus
TE	Time to Echo
TI	Time From Inversion
TNM	Tumour - Nodes - Metastasis
TR	Time to Repetition
UCLH	University College London Hospital
UICC	Union Internationale Contre le Cancre
UK	United Kingdom
US	United States
USA	United States of America
USB	Universal Serial Vus
V	Verticalis
VHF	Visible Human Female
VHM	Visible Human Male
VHP	Visible Human Project
Voxel	Volume Element
WHO	World Health Organization

Acknowledgements

First and foremost, I want to thank my supervisor: Dr. Stephen Golding. Without his patience, time and good sense, I never would have made it this far. I also never would have made it to the Chalet, or La Jonction.

My sincerest thanks to the staff of the Department of Oral and Maxillofacial Surgery, both at the John Radcliffe Hospital and University College London Hospital. You were all invariably helpful, no matter how demanding the request. Thanks you to Mr. Steve Watt-Smith, Mr. Colin Hopper and Mr. Nicholas Kalavrezos for their time, wisdom and support.

I also need to thank my colleagues in the Oxford MRI department: Stuart Meeson, Steve Turnbull, Dermot Dobson and Chris Alvey. Your input and ideas were invaluable. We also had a lot of laughs! Dr. Kostas Pataridis, without your experience and advice, things would have been much more difficult.

Words can't express my gratitude to the Rhodes Trust, but I'll try my best. It's been an unbelievable four years. Thank you for the financial freedom to experience Oxford and finish my DPhil. Mary Eaton, Catherine King, Bob Wylie, John Gee, Colin Page and Sheila Partridge: you've all gone above and beyond. Sir Colin Lucas and Dr. Don Markwell: you've both been generous with your wisdom and unwavering in your support.

Many thanks to my family and friends back in Newfoundland, especially my Mom and Dad. I've missed you all.

Rikki, you were by my side through it all. For that I'm forever grateful.

1 Introduction

1.1 Motivation

In modern surgical oncology, diagnostic imaging has become entrenched as a complementary adjunct to treatment and prognosis. The use of various imaging modalities has given surgeons a new perspective with which to approach patient care. It is not surprising that clinical management and prognostic guidelines have evolved to incorporate imaging-based criteria.

The role of diagnostic imaging in oral oncology is an interesting paradox. On one hand, state of the art techniques in three-dimensional computed tomography (CT) reconstruction give the surgeon an exquisite view of the bony structures of the head and neck and are a vital component of surgical planning. On the other hand, TNM staging criteria remain dependent upon semi-subjective and highly surgeon-dependent techniques of clinical examination.

TNM staging criteria are used not only for prognosis, but also to guide important decisions in surgical management. Oral cancer has a notorious propensity to metastasize to cervical lymph nodes. However, lymphadenopathy remains occult in the majority of cases, evading clinical detection. Accordingly, elective neck dissection accompanies resection in all but the smallest tumours, increasing surgical morbidity and mortality. A large number of patients that undergo neck dissection have no evidence of cervical node metastasis on pathology.

The need for refinement of the TNM staging system, as well as the development of a decision rule for elective neck dissection, is currently a widely debated topic in the medical community. Much work has gone into the study of clinical and histopathological factors that are able to predict prognosis and direct surgical management. The diameter of invasion, a key component of TNM staging, is chief among them. However, a lack of consensus has prevented the oral oncology community from building the strong portfolio of evidence required to support changing the current TNM staging criteria. Several factors contribute to this lack of consensus. The role of statistical methodology has been suggested for similar problems in other disciplines, but has yet to be explored in the oral cancer literature.

Invasion of the extrinsic muscles of the tongue is a criterion for T4a classification in oral cavity carcinoma. However, the role of the extrinsic muscles are ill defined. In many cases, the extrinsic muscles of the tongue lie just beneath the surface mucosa and can be invaded by superficial tumours that do not warrant T4a classification.

In the midst of this controversy, diagnostic imaging remains underutilized. MRI-measured diameter, thickness and volume have the potential to offer information to aid surgeons as they evaluate the natural history and prognosis of oral carcinoma.

1.2 Scope and Significance

This thesis looks at the implications to disease staging, prognosis and natural history in oral cancer of data inherent in examination by magnetic resonance imaging (MRI). Special emphasis is put on the role of tumour morphometry, including volume, diameter and thickness of invasion.

The depth of tumour invasion as a predictor of survival and occult cervical lymph node metastasis (OCLNM) has been studied extensively in recent years. However, while there is a general consensus that depth of invasion has a role in prognosis and prediction of metastasis, considerable controversy exists as to the most appropriate "cut-point" for establishing low risk and high risk subpopulations. Though several explanations for this cut-point variability have been proposed, the role of statistical methodology has received little attention.

Additionally, the vast majority of studies of depth of invasion use the histopathological measurement of resected oral carcinoma specimens. While histopathological analysis can be considered the gold standard in this area, it is not without its limitations. Post-surgical histopathological measurement is of little use when planning surgical management. In this capacity, MRI has the potential to provide multidisciplinary oncology teams with meaningful information at a very early stage in the patient journey.

It is important to realize that the information provided by MRI is just one component of an

integrated approach to patient management that also includes clinical and histopathological elements. As such, this work also examines a variety of other factors that have shown promise in the prediction of survival and OCLNM. These factors are combined with MRI-derived data to create predictive models using a number of approaches, including conventional multivariate regression, artificial neural networks and recursive partitioning analysis. The approaches are compared and conclusions are drawn as to the utility of each with special emphasis on the ability to generalize to external data sets. A major goal of this work is to propose new criteria for the TNM staging and surgical management of oral cancer, while also emphasizing a more individualized approach to prognosis.

The measurement of tumour dimension is not the only component of TNM staging that remains contentious among oral surgeons. The extrinsic muscles of the tongue are a key component of the definition of a T4a tumour. However, the extrinsic muscles are poorly identified on both imaging and pathology. Using the Visible Human Female (VHF) dataset, the extrinsic muscles of the tongue have been identified and described. Measurements have been taken to determine the surface proximity of the extrinsic muscles, particularly the hyoglossus and styloglossus.

Previous work in the department led to the development of an electronic anatomic atlas of the extrinsic muscles of the tongue suitable for co-registration with MRI images of the oral cavity [1]. Building on this proof-of-principle work, the atlas is used to delineate the position of the extrinsic muscles and evaluate invasion by squamous cell carcinoma of the oral tongue. The role of extrinsic muscle invasion in TNM staging is evaluated and recommendations made.

The main objectives of this thesis therefore can be summarized as follows:

1. *To design a retrospective cohort study of oral cancer patients with the intent of building multi-factor models of prognosis and occult cervical lymph node metastasis.*

2. *To establish a comprehensive database, consisting of demographic, clinical and pathological data, of patients diagnosed with oral cancer who have undergone MRI for staging purposes and have been managed surgically.*

3. *To conduct a comprehensive literature review of current approaches to the prediction of prognosis and cervical lymph node metastasis.*

4. *To evaluate the considerable controversy associated with the establishment of a depth of invasion "cut-point" for survival and occult cervical lymph node metastasis. Special emphasis is put on the role of statistical methods in contributing to the controversy.*

5. *To use univariate statistical analysis to determine significant predictors of all-cause, disease-specific and disease-free survival, as well as occult cervical lymph node metastasis.*

6. *To integrate significant factors into multifactor predictive models using a variety of techniques including multivariate regression, feed-forward artificial neural networks and recursive partitioning.*

7. *To validate and compare the above predictive modeling techniques with emphasis on the ability to generalize to external datasets.*

8. *To measure and evaluate the proximity of the extrinsic muscles of the tongue to the lingual surface using the Visible Human Female dataset.*

9. *To evaluate the role of extrinsic muscle invasion in TNM staging criteria using the fusion of an anatomic tongue atlas with clinical MRI images.*

10. *To bring the previous nine objectives together to present an integrated approach to the implications to disease staging, prognosis and natural history in oral cancer of data inherent in examination by magnetic resonance imaging.*

1.3 Organization

The thesis begins with a general introduction to provide an overview of the motivations and goals of the work. This is followed by general background and literature review of squamous cell carcinoma of the oral cavity including relevant anatomy and pathophysiology, epidemiology, the natural history of disease, surgical management and prognosis. Imaging of oral cavity cancer is discussed with emphasis on magnetic resonance imaging.

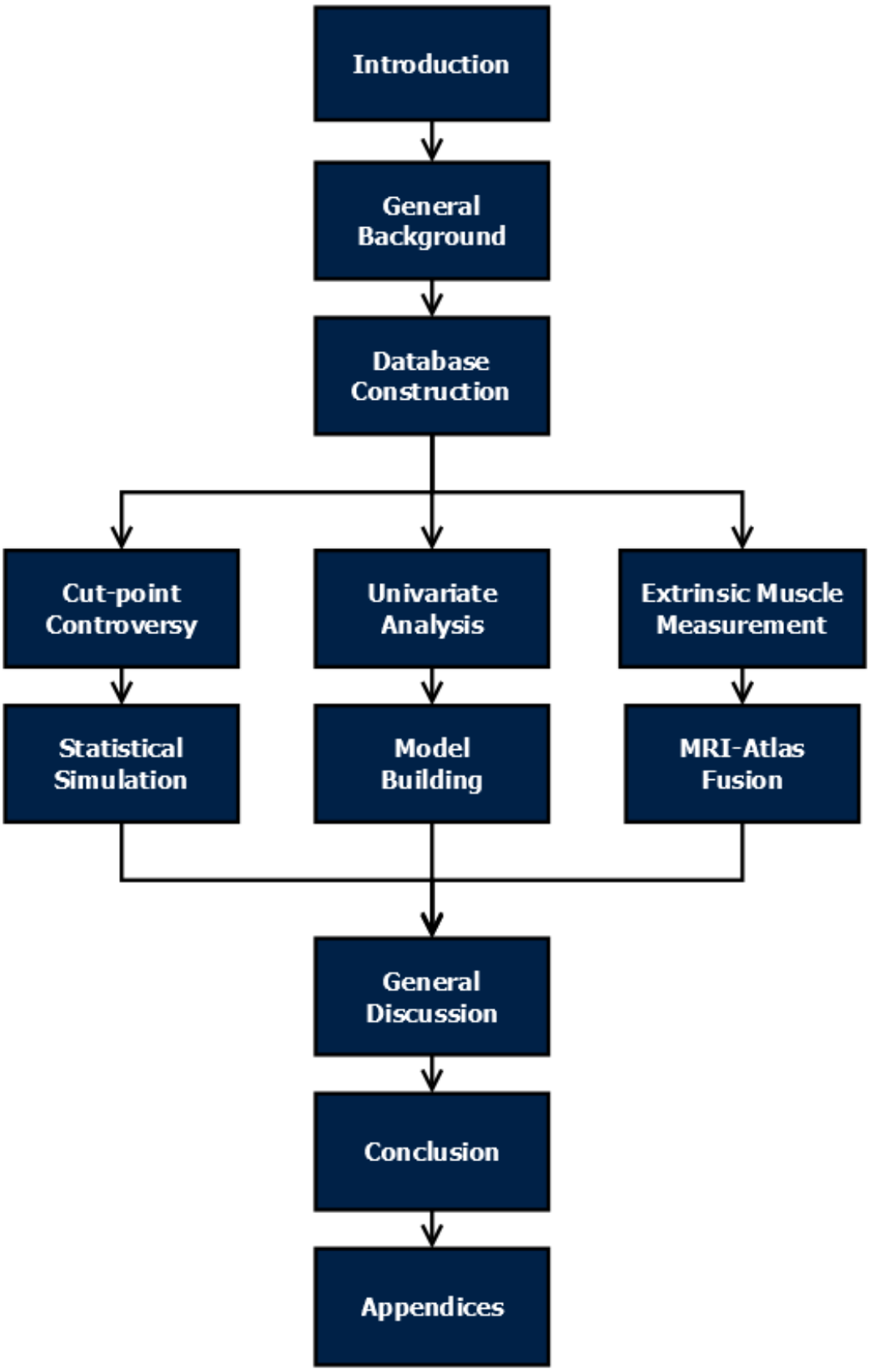
Database construction plays a role central to the components of this work. As such, associated methodology is discussed in detail and the content of the final database is outlined.

From this point onward, the work is best envisioned as following three parallel paths rather than a single linear progression. First, the controversy surrounding proposed cut-points for a TNM analogue for depth of invasion is outlined through comprehensive literature review. The contribution of statistical methodology to cut-point variability is considered through statistical simulation.

Second, each patient and tumour factor contained in the database is evaluated individually for the ability to predict occult cervical lymph node metastasis and survival through univariate statistical analysis. Explanatory variables identified as having predictive and prognostic potential are then used to construct multi-factor regression, artificial neural network and recursive partitioning models. The ability of these models to predict occult cervical lymph node metastasis and survival is compared with emphasis on their ability to generalize to external datasets.

Third, the role of the extrinsic muscles of the tongue in TNM staging is assessed. The extrinsic muscles are identified and described using the Visible Human Project Female dataset. The proximity of the extrinsic muscles to the surface of the lateral tongue is measured. An anatomic tongue atlas, developed in house, is fused to clinical staging MRI scans of oral cancer patients. The ability of the Atlas-MRI fusion to depict extrinsic muscle invasion is documented and the implications to TNM staging criteria considered.

The content of the thesis is then brought together in a general discussion. The work is summarized, common themes are outlined and proposals for further study are made. The thesis concludes with a brief summary and closing statement.


Introduction
General Background
Database Construction
Cut-point Controversy
Univariate Analysis
Extrinsic Muscle Measurement
Statistical Simulation
Model Building
MRI-Atlas Fusion
General Discussion
Conclusion
Appendices

2 General background

2.1 Introduction

This chapter begins with a brief description of the anatomy of the oral cavity and relevant adjacent structures. The text then moves on to discuss the epidemiology and histopathology of oral cavity cancer. The surgical management of oral cancer is covered including initial clinical examination, surgical resection, treatment of the cervical lymph node metastases and follow-up. Modalities for imaging oral cavity cancer are discussed with special emphasis on MRI.

From here the background shifts focus to the TNM staging of oral cancer. The principles of TNM staging as well as the TNM staging criteria for the oral cavity are outlined. This is followed by a discussion of the shortcomings of the TNM staging system, focusing on the inadequacy of superficial diameter. Alternatives to superficial diameter are explored with emphasis on the established role of the depth of invasion and the emerging role of the volume of invasion.

Finally, the rationale for inclusion in the database of the various prognostic and predictive factors is summarized. This is a key component of the background and sets the stage for the creation of multivariate predictive and prognostic models in Chapter 6. More detailed background is given, when necessary, in the relevant chapter. This general background is meant to serve as a description of topics that are common to all aspects of this thesis.

2.1.1 Chapter goals

The goals of this chapter are:

1. *To introduce the reader to relevant topics in oral cavity anatomy, magnetic resonance imaging, oral cavity cancer, surgical management and prognosis.*

2. *To provide the reader with a background for interpretation of subsequent chapters in this thesis.*

3. *To, where necessary, direct the interested reader to sources of further information for topics that are beyond the scope of this text.*

2.2 The anatomy of the oral cavity

In this section, the anatomy of the oral cavity is outlined, including the subsites, bony anatomy, vascular and nervous supply and lymphatic drainage.

2.2.1 Introduction

The anterior boundary of the oral cavity is defined by the vermillion border of the upper and lower lips. The posterior boundary is defined by the V-shaped terminal sulcus of the tongue, the junction of the hard and soft palates and the anterior border of the anterior faucal arch [2]. Each half of the oral cavity is divided into seven substructures, shown in Figure 2.1. The soft palate, the posterior one-third of the tongue and the fauces are considered part of the oropharynx [3].

2.2.2 The subsites of the oral cavity

2.2.2.1 The oral tongue

The tongue is divided anatomically by the V-shaped terminal sulcus. The anterior two-thirds of the tongue reside in the oral cavity. The pointed distal region of the body is termed the apex [2]. The surface anatomy is diagrammed in Figure 2.2. The complex muscular anatomy of the tongue is discussed in detail in Chapter 8 of this work.

2.2.2.2 The floor of the mouth

The floor of the mouth consists of the oral mucosa bordered by the oral tongue medially and the inferior alveolar ridge laterally and anteriorly. Its posterior extent is defined by the anterior tonsillar pillars. The mucosa of the floor of the mouth overlies the mylohyoid, hyoglossus and genioglossus muscles, which provide structural support [3].

2.2.2.3 The lips and cheeks

The external limit of the oral lip is defined by the vermillion border, the beginning of the transition zone between the external skin, which is heavily keratinized with hair follicles and sebaceous glands, and the less keratinized oral mucosa, which contains numerous minor salivary glands [3]. The cheeks and buccal mucosa are continuous with the lips and limit the oral cavity laterally.

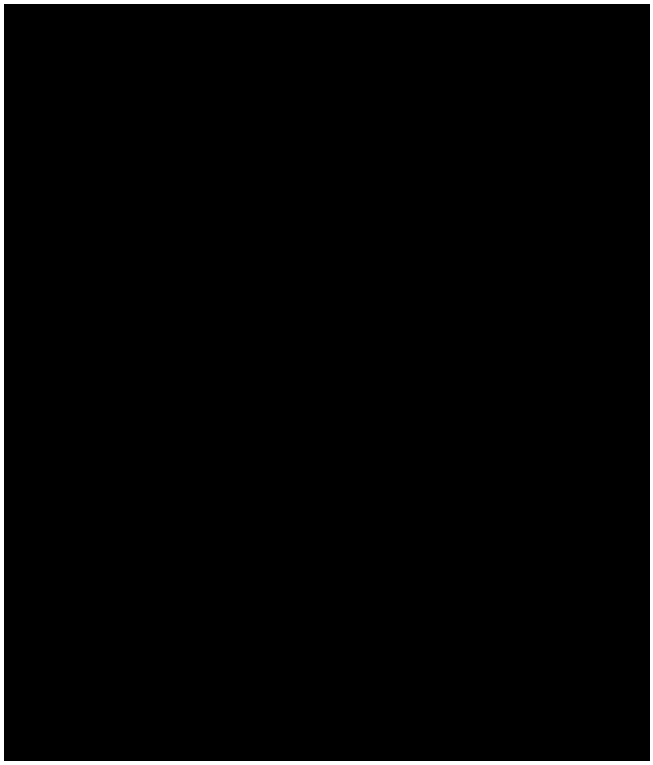


Figure 2.1: The subsites of the oral cavity. An anterior view of the oral cavity is shown and each of the subsites labeled. Source: [4].

Figure 2.2: The surface anatomy of the tongue. The oral tongue, also referred to as the body, is found anterior to the V-shaped terminal sulcus and comprises two thirds of the tongue. The root, or base, of the tongue is posterior to the terminal sulcus and is technically part of the oropharynx. Source: [2].

2.2.2.4 The lower gingiva and retromolar trigone

Overlying the alveolar processes of the mandible is the inferior alveolar mucosa. It is bound laterally by the transition to buccal mucosa and medially by the transition to the floor of the mouth. This thick, keratinized mucosa is adherent to the mandibular periosteum. The retromolar trigone mucosa is also adherent to the mandibular periosteum from the third molar to the maxillary tubercle. It is bound by the buccal mucosa laterally and the anterior pillar of the tonsils medially [3].

2.2.2.5 The upper gingiva and hard palate

The gingival mucosa overlying the maxillary alveolar process is adherent to the maxillary periosteum. The limits of the gingival mucosa are bound medially and laterally by the palatine mucosa and the buccal mucosa respectively. The mucosa overlying the hard palate is adherent to the palatial bone. The upper gingiva and hard palate are separated from the overlying maxillary sinuses and nasal cavity by only a very thin layer of bone [3].

2.2.3 The bony anatomy of the oral cavity

The bony anatomy of the face and skull base is complex and beyond the scope of this work. However, the reader is directed to Moore and Dalley for a more complete discussion [2]. The bony structures which contribute to the oral cavity include the mandible, the maxillae and the palatine bones, shown in Figure 2.3.

2.2.4 The vascular and nervous supply of the oral cavity

The lingual, facial and maxillary branches of the external carotid artery convey the blood supply to the oral cavity. Venous drainage enters the internal jugular vein [2].

Sensory innervation to the oral cavity is supplied by the maxillary and mandibular branches of the trigeminal nerve (CN V2 and CN V3). Motor innervation is by the facial nerve (CN VII) and the motor branches of the mandibular trigeminal nerve (CN V3). The intrinsic muscles of the tongue, the extrinsic muscles except the palatoglossus, and the geniohyoid receive motor innervation from the hypoglossal nerve (CN XII). The palatoglossus is supplied by the cranial root of the glossopharyngeal

nerve (CN XI). The vascular and nervous supplies of the oral cavity are diagrammed in Figure 2.4. For a more complete discussion the reader is directed to Moore and Dalley [2].

2.2.5 Lymphatic drainage system

The lymphatic network of the oral cavity communicates with the cervical lymph nodes. These basins contain anywhere from fifty to seventy lymph nodes and are divided into seven discrete levels by the anatomic structures of the neck [5]. The boundaries of each level are described in Figure 2.5. For a complete discussion of the anatomic boundaries of each level, the reader is pointed to Shah [5].

Levels six and seven are rarely involved in metastasis of cancers of the oral cavity and are not considered further. The lateral structures of the oral cavity tend to drain to ipsilateral cervical lymph nodes. However, it the lymphatics of midline structures have communications that cross the midline and drain bilaterally [3]. The most common routes for the lymphatic drainage of the oral cavity subsites is given in Table 2.1.

Table 2.1: Lymphatic drainage of the oral cavity subsites.

Oral Cavity Subsite	*Lymphatic Drainage Level*	*Contralateral Drainage*
Oral tongue	*I & II*	*Midline/Posterior*
Floor	*I & II*	*Common*
Upper alveolus	*I (anterior) & II (posterior)*	*Midline*
Lower alveolus	*I (anterior) & II (posterior)*	*Midline*
Retromolar trigone	*II*	*Rare*
Buccal mucosa	*IA & IB*	*Rare*
Hard palate	*II, lateral retropharyngeal*	*Midline*
Lower lip	*IA & IB*	*Midline*
Upper lip	*IA & IB*	*Midline*

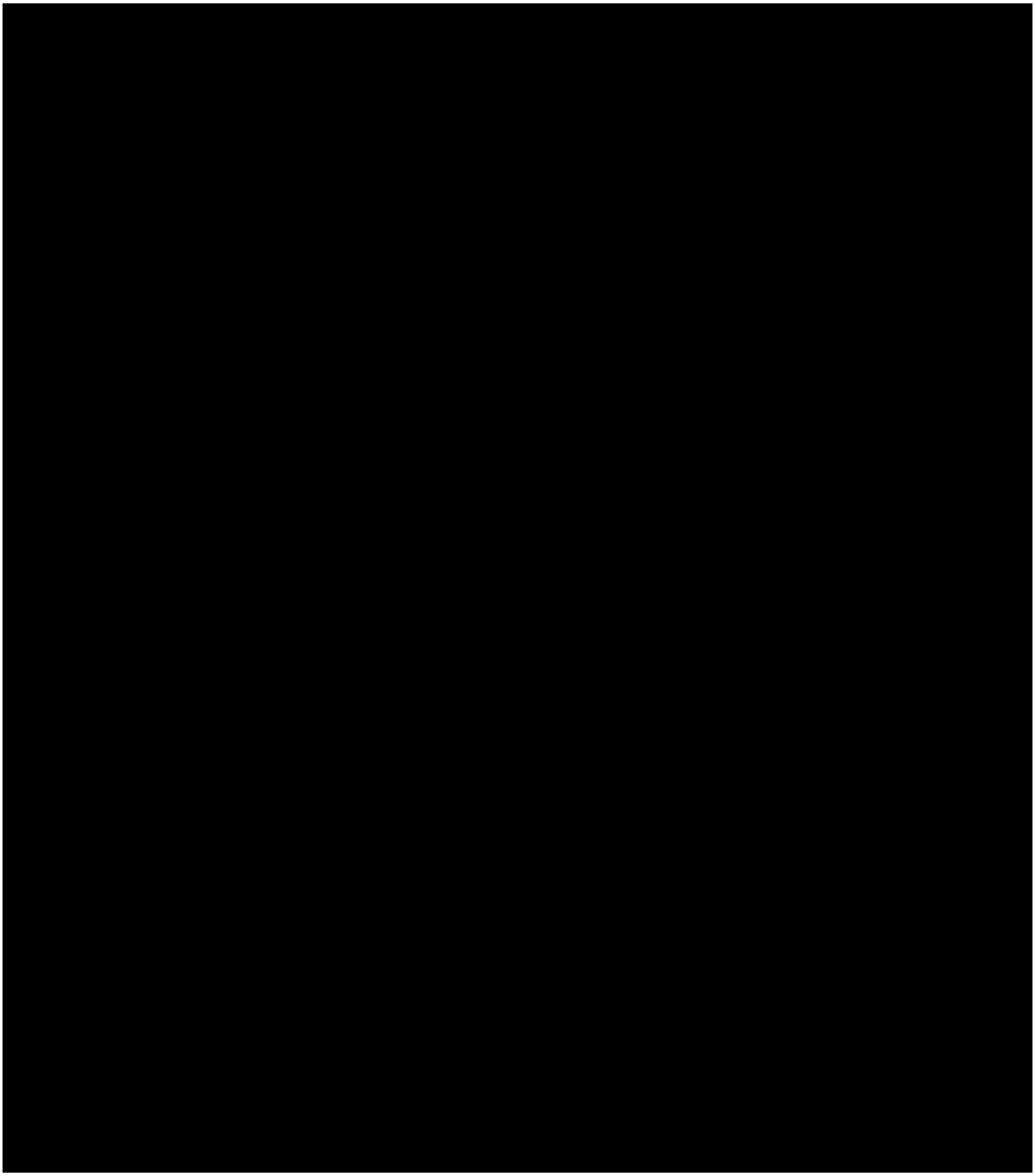


Figure 2.3: Bony anatomy of the oral cavity. The upper and middle left images depict anterolateral and posterior views of the mandible respectively. The lower left shows a medial view of the right hemi-maxilla. The right image shows an anterolateral view of the bony craniofacial anatomy. Source: [6]

Figure 2.4: The nervous and vascular structures of the oral cavity. A lateral view of the oral cavity with detailed illustration of the arterial blood supply and nervous distribution for the lateral oral cavity structures. Source: [6]

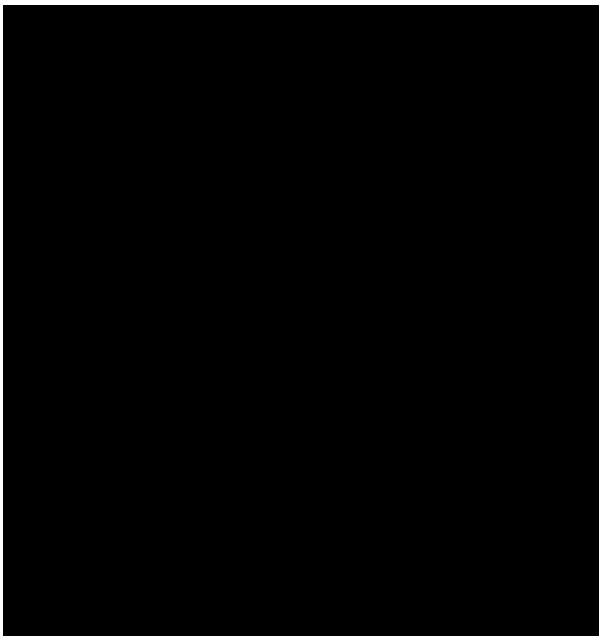


Figure 2.5: The cervical lymph node levels. Anatomical structures define the boundaries of discrete cervical lymph node levels. Source: [7]

2.3 MRI: A technical review

In this section, the principles of magnetic resonance imaging are reviewed. The concentration is on the technical aspects of MRI. MRI of the oral cavity is discussed in a future section.

2.3.1 Magnetic resonance imaging

MRI has risen to the forefront of diagnostic imaging buoyed by its exquisite soft tissue contrast resolution, multiplanar imaging and lack of ionizing radiation. Its success has been constrained by high costs, long image acquisition times, imaging artifacts and patient claustrophobia. Despite these limitations, MRI is the modality of choice for imaging the soft tissues of oral anatomy and pathology.

2.3.2 The basic principles of MRI

The spinning nucleus of the hydrogen atom has a small magnetic dipole moment. In a tissue sample, thermal energy randomizes the motion of the hydrogen dipoles and the net magnetic moment is zero. When an external magnetic field, B_0, is applied, the dipole moments of the individual protons align with B_0, either in the lower energy parallel state, or the higher energy antiparallel state (Figure 2.6). A slight excess of protons exist in the lower energy state producing a net magnetic moment [8].

Within an external magnetic field, the angular momentum of the proton spins causes them to precess about the axis of B_0 at an angular frequency, Ω_0, called the Larmor frequency:

Equation 2-1

$$\Omega_0 = \lambda \times B_0$$

where $\boldsymbol{\lambda}$ is the gyromagnetic ratio of hydrogen. Under normal circumstances, the individual protons precess with random phase and the net magnetic moment is parallel to B_0 [8].

Clinically, the patient is placed in the central bore of a large superconducting magnet such that B_0 is parallel to the cranial-caudal axis. Energy, in the form of radiofrequency (RF) electromagnetic waves, must be input into the system to extract information. The applied RF wave, precisely matched to the precessional frequency of the protons, imparts energy via resonant coupling.

Figure 2.6: Protons under the influence of an external magnetic field, B_0. At equilibrium, a there is a slight excess of protons with parallel spin. The input of radiofrequency energy at the Larmor frequency increases the proportion of protons with spins in the antiparallel state. Source: [9]

The RF pulse increases the proportion of protons with antiparallel spin, reducing and eventually inverting the longitudinal magnetic vector, M_Z. At the same time, the precession of the proton spins align and rotate in phase, creating a magnetic moment in the transverse plane, M_{XY}. The over-all effect is that the net magnetic vector is deflected away from the axis of B_0. The component of the net magnetic vector in the transverse plane can be measured using RF coils tuned to the Larmor frequency can be used to measure the magnitude of M_{XY}. This signal is derived from current induced in the RF coils by the rotating magnetic field [8].

2.3.3 Tissue dependent parameters: T1, T2 and proton density

The soft tissue discrimination of MRI is possible because of tissue specific variation in three key magnetic parameters.

2.3.3.1 T1 regrowth

Following excitation, protons in the antiparallel state lose energy to the environment as heat via spin-lattice interactions. The result is a return to the pre-RF equilibrium and the regrowth of longitudinal magnetization. The rate of regrowth is tissue dependent and can be used to distinguish tissues with differing physical properties. Defined, T1 is the time necessary for longitudinal magnetization to recover to 63% ($1 - 1/e$) of its equilibrium value [8], diagrammed in Figure 2.7.

2.3.3.2 T2 decay

RF application causes precessing protons to rotate in phase, generating a net transverse magnetic vector, M_{XY}. When the RF field is removed, ''spin-spin'' interactions cause small local variations in Larmor frequency and the spins dephase. The result is decay of transverse magnetization at a rate that is tissue dependent. Defined, T2 is the time required for transverse magnetization to decay to 37% ($1/e$) of its maximum post-RF magnitude. In the presence of external magnetic field inhomogeneities, decay of transverse magnetization is further enhanced, and the T2 parameter is referred to as T2* [8], diagrammed in Figure 2.8.

Figure 2.7: Tissue dependent parameter: T1. Following excitation, the protons return to the equilibrium energy state at a tissue dependent rate. T1 is defined as the time necessary for a tissue to recover 63% $(1 - 1/e)$ of its equilibrium value. Source: [9]

Figure 2.8: Tissue dependent parameter: T2. Following removal of the RF energy pulse, dephasing of the protons spins occurs at a tissue dependent rate. T2 is defined as the time required for transverse magnetization to decay to 37% $(1/e)$ of its maximum post-RF magnitude. Source: [9]

2.3.3.3 Proton density

The proton density refers to the number of proton spins per unit volume. Tissues differ in proton density. For example, fatty and aqueous tissues are rich in protons compared with proteinaceous tissue, or cortical bone [8].

2.3.4 The spin-echo pulse sequence (SE)

It is useful to consider the simple spin-echo (SE) pulse sequence, shown in Figure 2.9, to illustrate the timing parameters TE (time to echo) and TR (time to repetition). Together, the TE and TR parameters determine signal and contrast, giving an MRI image T1, T2 or proton density weighting.

The SE pulse sequence uses a 90° excitation pulse to ''flip'' the net magnetic vector into the *XY* plane, such that the magnitude of longitudinal magnetic vector, M_Z, is zero, and the magnitude of the horizontal magnetic vector, M_{XY}, is maximal. Spin-spin interactions lead to dephasing of the precessing protons. A 180° RF pulse is then applied, moving the net magnetic vector through 180°. This leads to rephrasing of the now out-of-phase protons. When the magnetic moments of the precessing protons are once again completely in phase, M_{XY}, and thus the signal induced in the RF receiver coil, reaches a maximum [10].

2.3.4.1 Time to echo (TE)

TE refers to the time from the initial 90° excitation pulse to the echo signal. TE determines T2 weighting. A short TE leaves little time for dephasing of the proton spins, thus T2 decay differences between tissues are minimal. Long TE allows more dephasing and the echo strength becomes more tissue T2 dependent [10].

2.3.4.2 Time to repetition (TR)

The TR refers to the time from the initial excitation pulse to start of the next pulse sequence. TR determines T1 weighting. Long TR allows complete recovery of M_Z following excitation and minimizes tissue T1 differences. Short TR thus emphasizes tissue T1 differences [10].

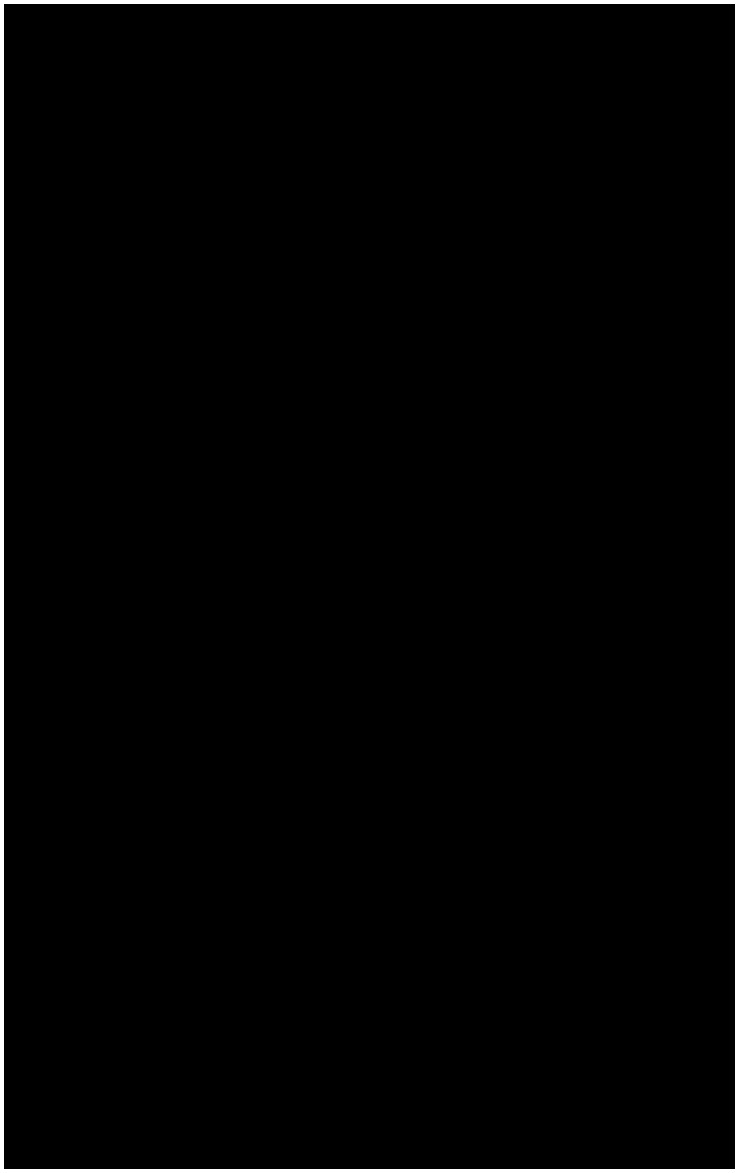


Figure 2.9: A generic spin-echo pulse sequence. A 90° excitation pulse ''flips'' the net magnetic vector into the XY plane, such that the magnitude of longitudinal magnetic vector, M_Z, is zero, and the magnitude of the horizontal magnetic vector, M_{XY}, is maximal. Spin-spin interactions lead to dephasing of the precessing protons. A 180° RF pulse is then applied, moving the net magnetic vector through 180°. This leads to rephrasing of the now out-of-phase protons. When the magnetic moments of the precessing protons are once again completely in phase, M_{XY}, and thus the signal induced in the RF receiver coil, reaches a maximum. RF: radiofrequency; Gs: slice selection gradient; Gp: phase encoding gradient; Gf: frequency encoding gradient. Source: [11]

2.3.4.3 Signal and contrast in spin-echo imaging

Together the tissue dependent parameters T1, T2, proton density and flow, and the pulse sequence parameters, TE and TR, determine the signal obtained from a tissue during a spin-echo MRI study:

Equation 2-2

$$S \; \alpha \; \rho_H f_{(V)} \left(1 - 2\varepsilon - \frac{TR}{T1}\varepsilon - \frac{TE}{T2}\right)$$

where S is signal, ρ_H is proton density and $f_{(V)}$ is flow. The contribution of flow is beyond the scope of this text and is not discussed. T1, T2 and proton density contribute to each and every MRI image produced. However contrast, the relative difference in intensity between different tissues, is determined by the pulse sequence parameters TR and TE. By varying TR and TE, the radiographer can produce images in which T1, T2 or proton density differences are accentuated [10].

2.3.4.4 T1 weighting

T1-weighted images are acquired using short TE and short TR. T1 weighting provides high signal from fat and low signal from water. Substances that order the random motions of water molecules, like gadolinium contrast agents or methaemoglobin, are also bright on T1 studies [10].

2.3.4.5 T2 weighting

T2-weighted images use long TE and long TR. T2 provides high signal from water and low signal from fat. Because magnetic field inhomogeneities hasten T2 decay, weakly paramagnetic substances, such as hemosiderin, appear dark [10].

2.3.4.6 Proton density weighting

Proton density-weighted images use short TE and long TR to minimize tissue differences in T1 and T2 constants. The resulting image contrast is dependent on the tissue proton density. Brain has a high proton density, while bone and air spaces have fewer protons [10].

2.3.4.7 Fast spin-echo (FSE)

In SE sequences a single 180° pulse is used per TR interval. Thus, only one line of k-space is filled per TR interval per slice. FSE uses a 90° excitation pulse followed by multiple 180° inverting pulses. This allows multiple dephasings and rephasings of the transverse magnetic vector, or multiple echos, per TR interval. By varying the phase encoding gradient with each inverting pulse, multiple lines of k-space are filled with each TR interval, decreasing the image acquisition time by a factor equal to the number of echoes per interval. Depending on the manufacturer, this factor is referred to as the echo train length or the turbo factor [10]. The principles of FSE and SE are similar, and T1, T2 or proton weighting may be produced. T2 images commonly take advantage of the speed of FSE.

2.3.4.8 Fat saturation and fat suppression

Fat is seldom considered an abnormal tissue while pathological tissue often has high water content, thus it is sometimes desirable to eliminate fat signal from an image. In addition, FSE images can often return high fat signal despite T2 weighting. This makes fat signal reduction a practical adjunct to T2-weighted FSE.

There are several techniques to minimize the signal intensity of fatty tissue. One method involves chemical saturation. In a 1.5 T superconducting magnet, differences in the chemical environments of fat and water hydrogens lead to a difference in precessional frequency of approximately 220 Hz. To saturate the fat signal a selective 90° pre-saturation pulse is applied at the precessional frequency of the fat hydrogens. The pre-saturation pulse if followed by the normal FSE 90° excitation pulse, which inverts the net magnetic vector of fat nuclei. All subsequent pulses are 180°, meaning that the net magnetic vector of the fat nuclei never gains a transverse component and no signal is recovered. Water has been spared the pre-saturation pulse and produces an echo train in the usual fashion [10].

2.3.5 The inversion recovery pulse sequence

Inversion recovery a pulse sequence used to suppress signal associated with certain tissues. The 90° excitation pulse is preceded by a 180° inverting pulse applied to the tissue to be imaged. The net

longitudinal magnetic vector, M_Z, is allowed to recover for a period of time, TI (time from inversion), before the 90° excitation pulse is applied to bring M_Z into the transverse plane. If the TI is chosen such that M_Z is 0 when the 90° excitation pulse is applied, M_{XY}, and thus the signal generated, will be zero. The time required for M_Z to recover from full inversion is dependent on the tissue specific factor T1. If we choose TI such that the M_Z of fat hydrogens is zero when the 90° pulse is applied, the fat signal will be nulled, while the signal from hydrogens in other chemical environments will not [10].

STIR, or short tau inversion recovery, is an example of an inversion recovery pulse sequence. It is used in conjunction with T2-weighted sequences to reduce fat signal [10]. STIR is used to illustrate the concept of inversion recovery in Figure 2.10 and Figure 2.11.

2.3.6 Spatial localization of MRI signal and image formation

MRI has the ability to reconstruct images in any plane. However, MRI is a cross-sectional technique, and relies on precise spatial encoding of MRI signals in the X, Y and Z planes.

2.3.6.1 Slice selection gradient (SSG)

Slice selection is accomplished by applying a linear SSG along the Z-axis of the stationary magnetic field. Because the Larmor frequency is proportional to the strength of the magnetic field, the linear gradient creates a range of precessional frequencies along the Z-axis. A narrow bandwidth RF pulse excites only those protons along the Z-axis that have the correct precessional frequency. The thickness is determined by the strength of the SSG, and the RF pulse bandwidth [8].

2.3.6.2 Frequency encoding gradient (FEG)

The FEG is applied orthogonal to the SSG, by convention the X-axis of a body scan. The FEG is activated during the formation and decay of the signal echo, generating a range of precessional frequencies dependent on X-axis position. The result is a complex waveform generated by the RF receiver coil that is decoded into positional and amplitude data by Fourier transform [12].

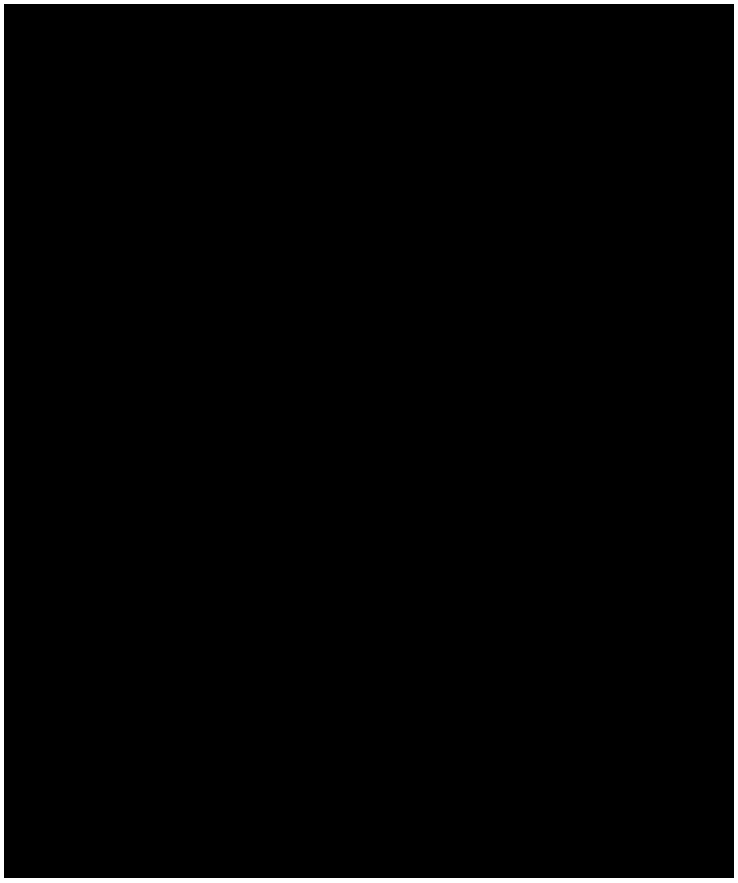


Figure 2.10: A short tau inversion recovery (STIR) pulse sequence. A 180° excitation pulse ''inverts'' the net magnetic vector, M_Z. M_Z is allowed to recover for a set time period, TI, the time from inversion. A 90° RF pulse is then applied, bringing M_Z into the *XY* plane. RF: radiofrequency; Gs: slice selection gradient; Gp: phase encoding gradient; Gf: frequency encoding gradient. Source: [11]

Figure 2.11: The time to inversion. If the TI is chosen such that M_Z is 0 when the 90° pulse is applied M_{XY}, and thus the signal generated, will be zero. M_Z recovery is a tissue dependent property, thus TI can be chosen such that the signal from fat hydrogens is nulled. Source: [11]n

2.3.6.3 Phase encoding gradient (PEG)

The spatial position in the axis orthogonal to the SSG and FEG is determined by the PEG. The PEG is applied after the SSG and before the FEG, and results in a linear variation in precessional phase throughout the gradient. Once the gradient is turned off, the precessional frequencies return to the Larmor frequency, but the phase variation remains. Each position along the PEG gradient axis is encoded with a specific phase shift [8].

2.3.6.4 Image reconstruction

For a given slice thickness, the PEG and FEG encoded data is digitized and mapped to k-space, the ''frequency domain'' repository. The filling of k-space can occur in a variety of ways, a discussion of which is beyond the scope of this text. k-space information is reconstructed in the spatial domain using the two-dimensional Fourier transform (2D Fourier) to extract information from phase, frequency and signal amplitude, forming the final image [8].

2.3.7 Characteristics of the MR image

The quality of the MRI image can be evaluated based on the contrast sensitivity, spatial resolution, and signal-to-noise ratio (SNR).

2.3.7.1 Contrast sensitivity

The contrast sensitivity refers to the ability to differentiate between two areas of differing tissue characteristics in an image. MRI soft tissue contrast is superb. Different tissues differ in their relative signal intensities which, in turn, depend on the weighting of the pulse sequence.

2.3.7.2 Spatial resolution

Spatial resolution refers to the ability to distinguish two adjacent objects as they become smaller and closer together and is determined by the field of view (FOV) and the size of the picture element (pixel) matrix. Together the FOV and pixel matrix size determine the pixel dimensions. As the pixel dimensions decrease, the spatial resolution increases, but not without trade-offs in other aspects of image quality.

2.3.7.3 Signal-to-noise ratio (SNR)

The SNR is a measure of the intrusion of background noise into the desired image signal and determines the ability to detect a low-contrast object in an image. The relationship between SNR and various factors is represented by the equation:

Equation 2-3

$$SNR\ \alpha\ I \times voxel_V \times \sqrt{NEX} \times \frac{1}{\sqrt{BW}} \times f_{(coil)} \times f_{(B)} \times f_{(slicegap)}$$

where I is the signal intensity, $voxel_V$ is the voxel volume, NEX is the number of excitations, BW is the receiver coil band width, $f_{(coil)}$ is the function of the coil selection and tuning, $f_{(B)}$ is the function of the magnetic field strength and $f_{(slicegap)}$ is the function of the slice gap effect [12].

The pixel dimensions, when multiplied by the slice thickness, yield the volume element (voxel) volume. The SNR is directly proportional to the voxel volume, thus the gain in spatial resolution by decreasing the FOV or increasing the pixel matrix size is offset by an increase in noise. However, large slice thickness can also result in partial volume averaging artefact and loss of contrast resolution. If there is no gap between slices, this can result in overlap and cross-excitation, reducing both SNR and image contrast [12].

If an area of the FOV is interrogated using multiple identical pulse sequence excitations, the data sets can be averaged, resulting in an increase in SNR, but also an increase in acquisition time. The RF coil must be selected such that it is as close to the volume of interest as possible. For example, the head coil is more appropriate for imaging of the oral cavity than a body coil. This close proximity, coupled with proper tuning to the Larmor frequency, maximizes detected signal and minimizes noise [12]. The final image quality requires careful consideration of all of these factors to achieve a balance between contrast sensitivity, spatial resolution and noise.

2.4 Oral cavity cancer

In this section, the nomenclature, epidemiology, etiology and histopathology of oral cavity cancer is

reviewed. The section concludes with background on local, regional and distant metastasis.

2.4.1 Nomenclature of carcinoma of the oral cavity

Studies of oral cavity cancer approach this heterogeneous area in a number of ways. The catchment area of oral and maxillofacial cancer surgery frequently includes oropharyngeal sites, thus many studies group the two together. The tongue is one area in particular where the distinction between oral and oropharyngeal cancers is often ignored. Technically, the oral cavity includes only cancer of the anterior two-thirds of the tongue. Studies involving the oral cavity and oropharynx offer a more complete sampling of lingual disease. Other studies distinguish between intraoral disease and cancer of the lip. Still wider classification systems group oral malignancy as cancer of the upper aerodigestive tract or cancer of the head and neck [3].

Table 2.2: Classification of subsites in the head and neck. Source: [13]

C00 - C14	*Malignant neoplasms of the lip, oral cavity and pharynx*
C00	*Lip*
C01	*Base of Tongue*
C02	*Other and unspecified parts of the tongue*
C03	*Gum*
C04	*Floor of the Mouth*
C05	*Palate*
C06	*Other and unspecified parts of the mouth*
C07	*Parotid gland*
C08	*Other and unspecified major salivary glands*
C09	*Tonsil*
C10	*Oropharynx*
C11	*Nasopharynx*
C12	*Pyriform sinus*
C13	*Hypopharynx*
C14	*Other and ill-defined sites of the lip, oral cavity and pharynx.*

Further distinction is made between cancers of other structures present in the oral cavity. The majority of oral cancers involve the mucosa. However, a significant number of malignancies arise in the sublingual, submandibular, parotid and minor salivary glands. Because the etiology and natural history of these cancers is markedly different, studies are often restricted to disease of the mucosa.

Care must be taken when interpreting the literature to be clear whether or not the authors are including or excluding disease of the salivary glands.

2.4.2 Epidemiology of oral cancer

Oral cancer is the 8th most common cancer worldwide amongst males and the 14th most common amongst women [14]. Globally, there are approximately 275 000 cases reported per year and 125 000 related deaths. In the developed world, oral cancer is the 11th most common amongst males and 18th amongst females. This is compared to the less developed world where oral cancer is ranked 7th and 10th amongst males and females respectively.

Figure 2.12 and Figure 2.13 show the global distribution of oral cancer in males and females respectively. The highest incidence of oral cancer is found amongst males in Papua-New Guinea at 40.9 per 100 000 (Age-Standardized Rate, ASR). The lowest ASR is found amongst females in Egypt at 0.2 per 100 000.

In the UK, a 2002 Globocan report estimates there were 3650 new cases of oral cancer and 1191 deaths. 60% of these new cases were male and 40% female making oral cancer the 13th and 17th most common malignancies [14]. The age-standardized 5-year survival for cancer of the oral cavity (excluding the oral tongue and lip) was 47.2% for males and 55.4% for females [15]. The figures for cancer of the tongue (including oral and oropharyngeal) were slightly less at 44.2% and 54.7% respectively. The 5-year survival for cancer of the lip was remarkably high at 94.9% and 91.1% respectively. Survival is shown according to gender in Figure 2.14 and Figure 2.15.

Oral cancer survival in the UK is age dependent. Figure 2.16 shows there is a 8 - 15% decline in survival when patients are grouped by increasing age [15]. Figure 2.17 shows that there is a gap of 16.3% and 11.6% in survival between the most affluent and least affluent groups for tongue and oral cancer respectively [15].

Figure 2.12: GLOBOCAN 2002 global oral cancer incidence (male). Areas of low incidence are shown in shades of green while areas of high incidence are shown in shades of red. Intermediate colours indicate moderate incidence. Source: [14]

Figure 2.13: GLOBOCAN 2002 global oral cancer incidence (female). Areas of low incidence are shown in shades of green while areas of high incidence are shown in shades of red. Intermediate colours indicate moderate incidence. Source: [14]

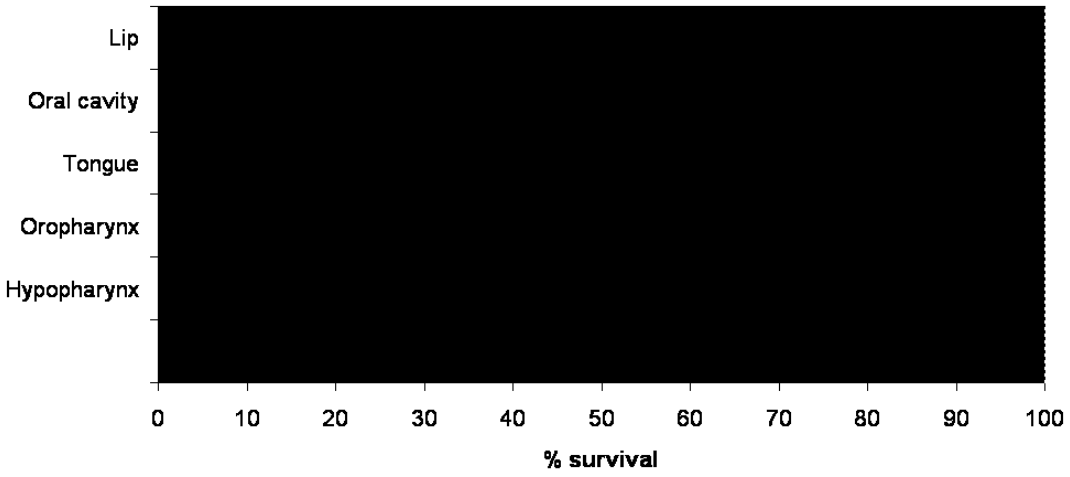


Figure 2.14: National statistics oral cancer survival (female). The dark blue bar indicates the period from 1996 to 1999 while the light blue bar indicates the period from 1991 to 1995. Survival improved for tongue and oropharyngeal cancers, but worsened for lip, oral cavity and hypopharyngeal tumours. Source: [15]

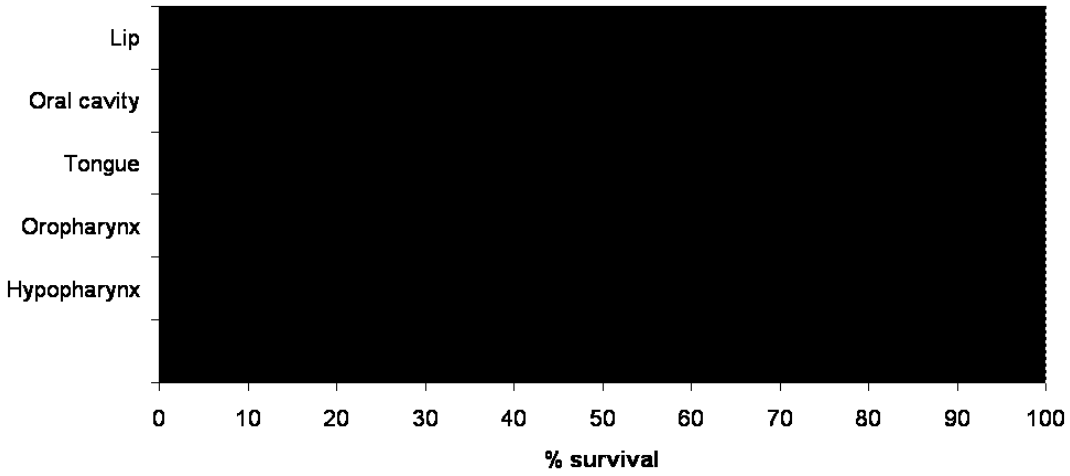


Figure 2.15: National statistics oral cancer survival (male) The dark blue bar indicates the period from 1996 to 1999 while the light blue bar indicates the period from 1991 to 1995. Survival improved for all tumours. Source: [15]

Figure 2.16: Age and socioeconomic status affect survival. All-cause survival rates decreased with increasing age for all tumours and patient genders. Survival was also slightly higher in females compared to males. Source: [15]

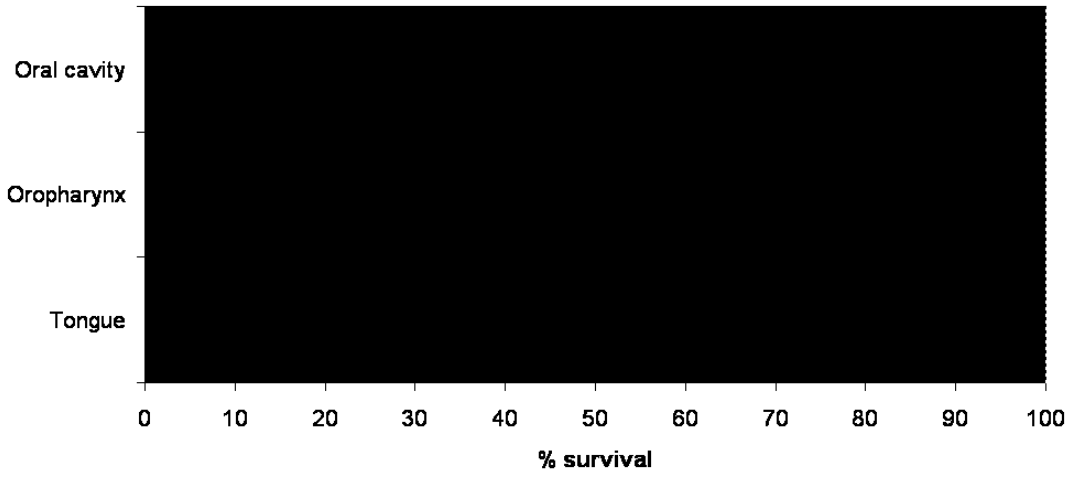


Figure 2.17: Age and socioeconomic status affect survival. Those patients who were the most deprived (dark blue) had a substantially reduced survival percentage compared to the most affluent (light blue). The trend was found in all three tumour groups. Source: [15]

2.4.3 Etiology of oral cancer

2.4.3.1 Alcohol and Tobacco

Together, alcohol and tobacco are associated with more than 75% of oral cancers [16]. Smokeless tobacco, as ''chew'' or ''snuff'', contains a variety of carcinogens and is also associated with oral cancer [17].

Alcohol is the second largest risk factor for oral cancer [18]. This results from direct effects of ethanol on oral mucosa, as well as systemic effects of chronic ethanol abuse and associated nutritional deficiencies [19]. Together, alcohol and tobacco act synergistically. Their effects on cancer development have been shown to be multiplicative rather than additive [20].

2.4.3.2 Oncogenic viruses

Several viral agents have been implicated in the genesis of oral cancers including human papilloma virus (HPV), Epstein-Barr virus (EBV) and hepatitis C virus (HCV).

HPV has been implicated in a number of human cancers, most notably cervical cancer. The HPV viruses are a heterogeneous group of over one hundred known serotypes with varying oncogenic potential. The reported incidence of HPV infection in oral cancers varies widely, mostly because of the variable sensitivity and specificity of laboratory detection. Recent studies report the incidence to be as high as 91% in precancerous and cancerous lesions [21]. It is likely that HPV acts synergistically with chemical carcinogens in alcohol and tobacco. The role of HPV varies with anatomic location, with HPV playing a smaller part in oral cancers compared to oropharyngeal cancers. Patients with HPV-positive oral SCCs tend to be younger, drink less alcohol and have a better prognosis.

The evidence for EBV and HCV is less convincing. EBV has been detected more frequently in patients with oral lichen planus or oral SCC than in healthy mucosa [22-23]. However, several conflicting studies exist and further work is required [24-26]. Individuals with chronic HCV infection have an increased risk of oral SCC, implying that the HCV virus plays some yet to be elucidated role [27-28].

2.4.3.3 Genetic predisposition and diet

There is some evidence that there is an increased incidence of malignancy in first degree relatives of head and neck cancer patients, with reported relative risks of 1.1, 2.5, 3.5 and 3.8 [29-32]. As well, oral cancer patients with first degree relatives with upper aerodigestive tract malignancies are at increased risk for second primary tumours, indicating enhanced field cancerization [33]. A very small proportion of oral cancer patients may display an autosomal dominant inheritance pattern [34].

A study by La Vecchia et al. estimates that in Europe, as many as 15% of oral and oropharyngeal cancers are attributable to dietary deficiencies [35]. The group showed that green vegetables and fresh fruits are protective with odds ratios of 0.6 and 0.2 respectively. Similarly, a Beijing study demonstrated a protective effect from carotenoids and vitamin C, dietary fiber [36].

2.4.4 Histopathology of SCC

2.4.4.1 Pre-malignant lesions

Leukoplakia is defined as a white patch or plaque of oral mucosa that cannot be removed by vigorous scraping [37]. Leukoplakia progresses to invasive SCC in approximately 6% of cases, while regressing or remaining unchanged in most [38].

Erythroplakia is strictly defined as a velvety red patch that cannot be ascribed to any specific disease process [37]. It is thought to be the first sign of asymptomatic invasive SCC and has a rate of malignant transformation approaching 50% [38]. Lesions often consist of both leukoplakia and erythroplakia, referred to as speckled leukoplakia. Erythroplakia and leukoplakia are part of a continuum oral intraepithelial dysplasia (OIN) with progressively increasing clinical and pathological alterations, shown in Figure 2.18.

Microscopically, oral dysplasia is characterized as mild, moderate or severe depending of the extent of atypical cellular features. *Carcinoma in situ* is characterized as the involvement of the full thickness of the epithelium by dysplastic cells in the presence of an intact basement membrane. In this way, the classification of OIN is similar to cervical intraepithelial dysplasia (CIN) [37].

2.4.4.2 Malignant lesions

Clinically, SCC presents as a tumor mass, an ulcer, or a region of leukoplakia or erythroplakia. Histologically, the basement membrane is breached and the neoplasm extends into the dermis as broad sheets, cords, nests and islands of neoplastic cells.

Neoplasms are classified subjectively as well, moderately or poorly differentiated based on microscopic similarity to the native epithelium, illustrated in Figure 2.19 [37]. Well differentiated tumours are composed of cells that are similar in appearance to native epithelium. Atypical cellular features are minimal and keratin formation is commonly found. Keratinization is comparatively rare in moderately and poorly differentiated SCC [37]. Poorly differentiated tumours bear little, if any, resemblance to normal oral mucosa. The frequency of atypical cellular features is greatly altered with hyperchromatism, pleomorphism, anisocytosis, loss of intercellular cellular attachments and atypical mitoses common features [37]. Moderately differentiated tumours may still bear some resemblance to native epithelial cells, but atypical cellular features are present to a greater degree. A moderately differentiated tumour is best described as a tumour that can be described as neither well nor poorly differentiated.

2.4.5 Local, regional and distant spread of SCC

2.4.5.1 Local spread

Oral cancer follows two distinct patterns of invasion. When invasion is cohesive, the leading edge of the tumour remains well defined. Non-cohesive infiltrative growth occurs when the tumour's leading edge is poorly defined. Islands or tendrils of malignant cells can be seen at the infiltrative front and are associated with increased risk of local recurrence and regional metastasis [39-40].

Because of extensive heterogeneity among oral cavity subsites, the pattern of local spread is often site-specific and dependent on surrounding tissues. In the oral tongue, advanced cancers cross the midline and invade the extrinsic muscles. The extrinsic muscles serve as conduits for spread to the remainder of the oral cavity and oropharynx [41].

Figure 2.18: The spectrum of intraoral epithelial dysplasia. Oral dysplasia can be thought of as a spectrum from normal mucosa, through various forms of leukoplakia to erythroplakia. Each point on the spectrum is associated with increasingly clinical and pathological alterations. Source: [37]

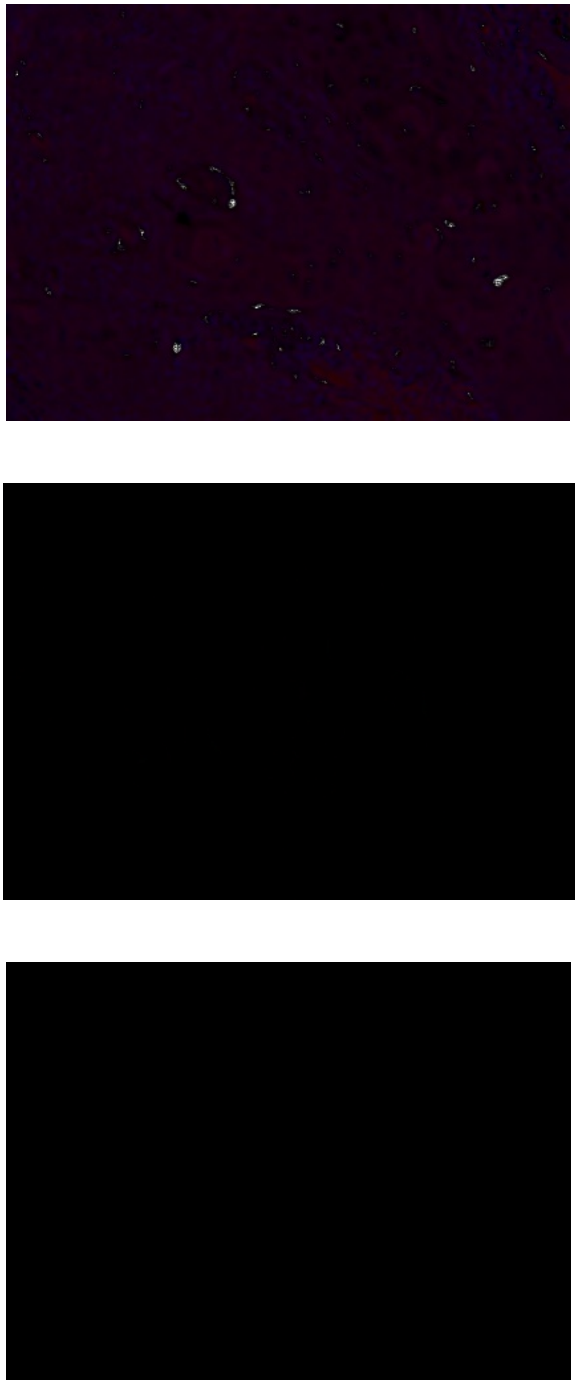


Figure 2.19: Various levels of differentiation of oral cavity squamous cell carcinoma. The top shows a well differentiated SCC demonstrating keratin pearl formation (arrow). The middle shows a moderately differentiated SCC with accumulations of cells somewhat identifiable as squamous in origin, as well as some keratin. The bottom photograph shows a poorly differentiated SCC with aggregations of tumour cells and no keratin. Source: [37]

Floor of the mouth tumours are adjacent to a number of structures. Superior spread is to the ventral tongue, while lateral and anterior spread may involve the mandible. Inferiorly, the genioglossus and mylohyoid muscles can be invaded [42].

Tumours of the upper and lower gingiva, as well as the retromolar trigone, are well placed to invade underlying bone. Tumours of the buccal surface of the gingiva may reach the cheek, while lingual surface tumours may extend to the hard palate or floor of the mouth [42].

Buccal tumours extend along the buccal mucosa and can involve the full thickness of the cheek. Large tumours can extend to the buccal surface of the upper and lower gingiva [42]. Hard palate tumours closely approximate the maxilla. Bony erosion and extension into the maxillary or nasal cavity are common. Large tumours may extend to the lingual surface of the upper gingiva [42].

2.4.5.2 Regional metastasis

Lymphatic metastasis is a primary concern of the oral surgeon. Nodal metastasis is associated with a 50% decrease in five year survival [43]. Occult metastasis is found on pathology in 30-40% of individuals with negative clinical and MRI examinations of the neck [44-47].

In a seminal study by Shah of oral cancer patients undergoing 192 elective radical neck dissections, 146 lymph nodes positive for metastasis were found [48]. 58 of these were in level I while 52 were in level II. 26 involved level III nodes while only 9 and 2 metastatic nodes were found in levels IV and V respectively. The frequency of level involvement is also determined by oral cavity subsite and is known to reflect the lymphatic drainage routes outlined in Section 2.2.4.

The pattern of metastasis is not always predictable. In a study of a large cohort of lingual cancer patients by Woolgar [49], cases of skip metastasis were noted, in which affected nodes are found in non-contiguous neck levels. ''Peppering'' was also documented, in which metastases were found at multiple levels of the neck without an obvious primary focus.

Once lymphatic metastasis has become established, it may invade beyond the lymph node border.

This extracapsular spread (ECS) is associated with decreased survival [50].

2.4.5.3 Distant metastasis

Distant metastasis occurs late and follows the lymphatic or hematogenous route. Common sites include bone, brain, liver and lung. The incidence of distant metastasis is high in local cancer, with rates approaching 50% [51]. ECS is the most significant predictor of distant metastasis. 71% of patients with ECS were found to have distant spread [52].

2.5 Surgical management of lingual cancer

In this section, surgical management is outlined, including presentation, clinical examination, imaging, primary resection, neck dissection and defect reconstruction.

2.5.1 Presentation

Oral cancer lesions may be discovered as an incidental finding by the general practitioner or dentist. Alternatively, oral cancer may arise from an area of oral dysplasia that has been monitored for an extended period. Lesions may present under a variety of guises. The patient may notice an ulcer or exophytic growth that is enlarging rapidly, or fails to heal. Such lesions may bleed, be painful, result in poorly fitting dental appliances, trismus, dysphagia or halitosis. Otalgia or paresthesia may result from nerve invasion [53]. The index of suspicion is raised by the presence of risk factors such as a history of smoking, alcohol consumption or male sex [53].

2.5.2 Pre-operative clinical examination

Bimanual palpation is used to identify areas of tissue that are tender, indurated or fixed [54]. Examination of the submental, submandibular and cervical lymph nodes may reveal lymphadenopathy. 25-35% of oral SCC has palpable submandibular or cervical lymph nodes at presentation [55]. Suspicious cervical lymph nodes may also be sampled with fine needle aspiration (FNA) at some institutions. In cases of SCC of the posterior oral cavity, panendoscopy might be necessary to explore the upper aerodigestive tract. Complete examination is important as rates of

synchronous lesions have been reported to be 5-10% [53].

The tumour is biopsied as part of pre-operative evaluation. Biopsy is carried out at the perimeter of the lesion and sent to a histopathologist for tissue diagnosis.

2.5.3 Pre-operative imaging

Pre-operative imaging is performed to assess local, regional and distant tumour spread. The primary lesion may be assessed with either CT or MRI, depending on institutional preference and availability. MRI provides excellent soft tissue contrast and overcomes problems associated with dental artefact that frequently obscures CT studies. PET-CT may play a role in situations where the primary lesion is unidentified. When bony invasion is suspected, CT or orthopantomogram (OPG) is used to assess cortical bone erosion. MRI has a limited role in the evaluation of bony invasion but is able to identify medullary invasion.

In the neck, the tomographic modality used to evaluate the primary lesion is extended to the clavicles to allow evaluation of cervical lymph node metastasis. Rarely, PET-CT is used to assess the neck. Chest radiography or chest CT is used to screen for distant lung metastases.

2.5.4 Tumour resection

The tumour is resected with a wide margin between 1 cm and 2 cm. Approaches to surgical access are shown in Figure 2.20 and Figure 2.21.

2.5.4.1 Oral tongue

Types of glossectomy range from partial for a small lesion to total for extensive tumours. As the extent of glossectomy increases, so does the degree of speech and swallowing impairment, decline in quality of life, and incidence of recurrence [4].

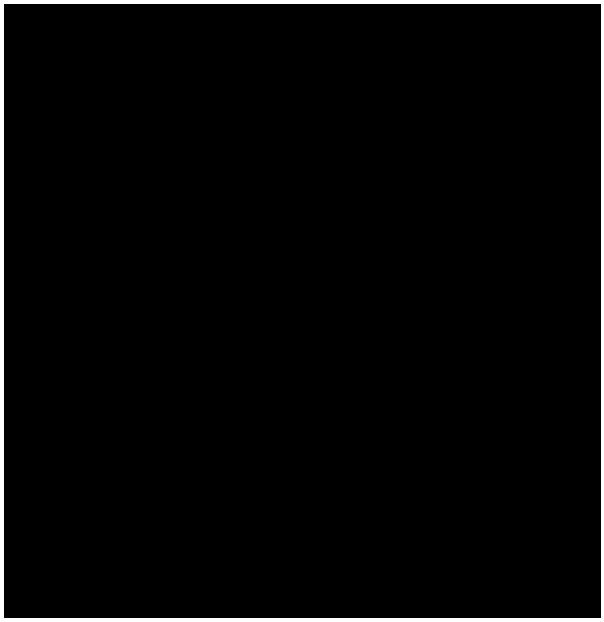


Figure 2.20: Stair step mandibulotomy. The lateral portion of the mandible is swung away from the oral cavity, exposing the tumour for resection. The stair step aides realignment. Source: [4]

Figure 2.21: The visor flap. The upper illustration depicts the incision from mastoid to mastoid. In the lower illustration, the resulting skin flap is retracted superiorly to expose the oral cavity and mandible. Note the exposed mental foramen. The mental nerves have been divided. Source: [4]

2.5.4.2 Floor

Lesions of the floor often invade the underlying muscles and may require resection of the genioglossus, hyoglossus and mylohyoid. Extension into the oral tongue may necessitate glossectomy in continuity with the primary lesion [53]. Extension to the mandible is an ominous sign that requires careful management, considered in Section 2.5.4.4.

2.5.4.3 Buccal Mucosa

Local excision is the treatment of choice for small lesions. Deep lesions may require resection of the buccinator muscle or a through and through resection. Extensive tumours may require resection of the mandibular or maxillary alveolar ridge [53].

2.5.4.4 Retromolar trigone, lower alveolus and the mandible

While small, localized lesions may be excised, retromolar trigone and lower alveolar lesions are notable for their close proximity to the mandible. Depending on the extent of invasion, a variety of approaches to the mandible may be considered [53].

For lesions that approximate, but do not invade, the periosteum, removal of the adjacent periosteum may be indicated. When periosteal invasion is present, a marginal mandibulectomy is indicated. In the coronal version of the marginal resection, the upper aspect of the involved mandible is removed. In the lingual sagittal version, the cortex of the lingual surface of the affected mandible is resected [53]. Extensive lesions require a segmental resection, disrupting the continuity of the mandible and necessitating advanced reconstruction procedures.

2.5.4.5 Hard palate, upper alveolus and the maxilla

Tumours of the hard palate and upper alveolus overlie the maxilla. Like lesions of the mandible, periosteum removal may be sufficient in cases with limited spread, however periosteal involvement requires bony resection. Partial or subtotal maxillectomy is required for large lesions [53].

2.5.5 Neck dissection

Neck dissection is a frequent adjunct to glossectomy. It is performed both when nodal disease is present and in the clinically negative neck [48]. The various approaches are listed in Table 2.3.

Radical neck dissection (RND) is reserved for patients with extensive node disease, extending into the spinal accessory nerve or internal jugular vein. Removal of the spinal accessory nerve results in denervation of the trapezius, causing pain, dysfunction and deformity [7]. Modified radical neck dissection (MRND) or functional neck dissection (FND) are used when there is confirmed lymph node disease that does not extensively invade surrounding structures. MRND is also commonly used to manage the clinically negative neck, especially when the primary tumour is extensive [7]. Selective neck dissection (SND) is recommended for the clinically negative neck in which there is a risk of occult metastasis [7].

Table 2.3: Surgical approaches to neck dissection in the oral cancer patient.

Approach	Structures Removed
Radical neck dissection	*All lymph nodes from level I-V, spinal accessory nerve, internal jugular vein, sternocleidomastoid muscle*
Modified radical neck dissection	*All lymph nodes from level I-V. One or more of the spinal accessory nerve, internal jugular vein or sternocleidomastoid are spared.*
Functional neck dissection	*All lymph nodes from level I-V. The spinal accessory nerve, internal jugular vein and sternocleidomastoid are all preserved*
Selective neck dissection	*One or more lymph node groups. For lingual malignancies this is typically levels I-IV. Also known as supraomohyoid dissection.*

The decision to offer neck dissection in the management of the negative neck is a contentious issue with prognostic significance [56-57]. The controversy extends to all aspects of the procedure, including the efficacy of investigations to determine the status of the cervical lymph nodes, the need to treat, the optimal treatment modality, the extent of therapy, and the necessity of contralateral neck management.

CT and MRI are currently used to evaluate the neck for metastases, but have a sensitivity somewhere between 40-60% [58]. The only role for observation is in small T1 carcinomas [59]. Radiation has

been shown equivalent to neck dissection in terms of survival and avoids surgical morbidity and mortality, but does not offer the prognostic information about the presence of occult metastasis [60]. Contralateral dissection is reserved for tumours that cross the midline, despite evidence that there are extensive contralateral lymphatic communications.

2.5.6 Reconstruction of the surgical defect

Small excisions may heal well by second intention or primary closure. Split thickness skin grafts may be required to close larger mucosal defects or cover exposed bone. A variety of advancement and rotational flats may be considered for more extensive defects. Options include pectoralis major and latissimus dorsi flaps [53].

Free tissue transfer is used extensively in the reconstruction of significant oral cavity defects. Microvascular anastamosis of flap vessels to local vessels is required and mandates special expertise. In some cases, nerves can be anastamosed and provide a sensate flap. A variety of free flaps are available and offer the surgeon a number of options in terms of mucosal surface area, fat and muscle bulk and vessel and nerve availability. These sites include the anterolateral thigh, forearm and rectus abdominus. When bone is required for mandibular reconstruction, flaps can be raised that include the fibula, iliac crest or scapula [53].

2.6 MRI of oral cavity carcinoma

A variety of techniques are available for the imaging of oral cavity carcinoma. A complete discussion of these techniques and their indications is beyond the scope of this work, however the reader is directed to Ahuja [61]. This work will concentrate on the techniques, indications, strengths and weaknesses of MRI.

2.6.1 MRI acquisition parameters

Dedicated head coils can offer superior resolution and SNR for imaging of the oral cavity; however coverage is limited to the oral cavity and skull base. For oncology applications, the cervical lymph

nodes are also of primary concern and neck imaging is required as part of comprehensive staging. In a busy imaging department it is time consuming and impractical to scan the head and neck separately [61]. Instead, a combined head and neck coil is used to cover the entire region of interest in a single set of scans. However, the heterogeneity of the surface geography of the head and neck lead to coil construction parameters that sacrifice some image quality for increased coverage.

Other parameters that are of vital importance to image quality are the field of view (FOV), the slice thickness, the inter-slice gap and the pixel matrix size. The FOV should be chosen such that the region of interest is imaged in its entirety. However, there is a compromise between FOV and SNR. A small FOV will maximize spatial resolution with a concomitant fall in SNR. When excessively large FOVs are used, SNR will improve, however the region of interest will occupy only a small portion of the image area. Most authors recommend the use of a FOV ranging from 14 X 14 cm to 16 X 16 cm for T1-weighted images and 18 X 18 cm to 20 X 20 cm for T2-weighted images as a suitable compromise. The choice of the slice thickness must balance a compromise between SNR and partial volume averaging. As slice thickness increases SNR also increases, however increased partial volume averaging leads to a loss of spatial and contrast resolution. The recommended slice thickness varies between 3-6 mm. Inter-slice spacing varies from 0-1 mm [61].

The pixel matrix size selection requires a compromise between SNR and spatial resolution. As the pixel matrix size increases, so does the spatial resolution. However there is a concomitant loss of SNR. 1.5T scanners use 512 X 512 matrix dimensions while 256 X 256 is still used for selected sequences [61]. Modern 3.0T scanners are not yet in wide use, but with all other parameters equal, the 3.0T scanner, theoretically, offers double the SNR of a 1.5T scanner. In practice, this allows radiographers to reduce scan times, reduce slice thickness, reduce field of view, increase pixel matrix size or some combination of the aforementioned parameters.

Table 2.4: Commonly used equipment and parameters for MRI studies of patients with oral cavity cancers. Adapted from Imaging in Head and Neck Cancer: A Practical Approach [61].

Parameter	*Technique*
MR Coil	*Dedicated combined head and neck coil* *Head coil, or neurovascular coil, to cover primary tumour and nodes from the skull base to the oral cavity* *Neck coil to cover primary tumor and nodes from the oral cavity to the suprasternal notch*
Scan Plane	*The coronal and axial planes are commonly used.* *Sagittal planes may be used as an additional plane to cover the skull base, tongue root, floor of the mouth and posterior hypopharynx*
Sequence	*T1-weighted spin echo or fast spin echo* *T2-weighted fast spin echo with or without fat suppression* *T1-weighted spin echo or fast spin echo post-contrast, with or without fat suppression*
Image Parameters	*Slice thickness 3-6 mm, with 0-1 mm spacing* *FOV as small as possible (14-16 cm for T1, 18-20 cm for T2)* *Matrix size preferably 512 with 256 for selected sequences*

2.6.2 Appearance of the oral cavity on MRI

Historically there has been more interest in MRI of the tongue relative to other oral cavity sites. Early work by Unger was able to identify the genioglossus and geniohyoid muscles [62]. Lufkin followed a year later in 1986 with improved technology, and was able to demonstrate the superior longitudinal, genioglossus and transverse muscles with increased spatial resolution [63].

T1-weighted sequences provide excellent anatomic detail of the soft tissues of the head and neck. Images are typically acquired using spin echo (SE) or fast spin echo (FSE) with short TE and TR. In many institutions, T1-weighted images may also be routinely acquired with the addition of a gadolinium contrast agent. T2 images are also typically acquired using FSE [64].

On T1-weighted images, the muscles of the oral cavity are identifiable as structures of uniform low signal. However, within the muscles of the oral cavity there can be variability in signal intensity due to variable fat content in the muscle substance. The extrinsic and intrinsic muscles of the tongue are usually distinguishable. Fat planes separating the oral cavity structures are also prominent features.

The fibrofatty lingual septum is visible in the midline sagittal plane. While the genioglossus' fat content is relatively constant, there is much more variability in the fat content of the intrinsic muscles. The muscles of the oral cavity are also of uniform low signal on T2-weighted images, however fat planes, such as the lingual septum, are no longer distinguishable [64].

The glandular structures of the oral cavity are of high signal intensity, relative to muscle, on both T1 and T2-weighted images. The same is also true of the lymphoid tissues, which are concentrated in Waldeyer's ring in the oropharynx [64].

Regardless of pulse sequence, cortical bone appears as low signal due to the low proton density of the mineral matrix. The soft tissue of medullary bone is distinguishable from that of cortical bone, particularly in the mandible. On T1 images the fatty marrow of medullary bone gives high signal while on T2, the marrow is of intermediate signal [61]. Neurovascular structures are sometimes visible on MRI of the oral cavity.

The use of gadolinium-based contrast materials is useful for highlighting neoplastic growths; however the signal characteristics of normal structures are also changed. The tissue of the oral mucosa is seen to enhance. There is also a modest enhancement in signal intensity of glandular and lymphatic structures. Other oral cavity structures remain relatively unchanged [64].

2.6.3 Appearance of oral carcinoma on MRI

2.6.3.1 T1-weighted images

Lufkin conducted the first work detailing oral cavity SCC in 1988 [65]. On T1-weighted images, SCC is isointense or has a low signal intensity compared with surrounding tissue. If the fat content of the muscle is high, as in the tongue, the tumour will be clearly visible. However, if fat content is low, contrast resolution between tumour and muscle will be less [66]. On all pulse sequences, tumour can also be inferred from mass effect distorting the oral cavity symmetry, especially the prominent lingual septum. An example is shown in Figure 2.22.

2.6.3.2 T2-weighted images

T2 weighting demonstrates superior contrast and is most sensitive to the presence of tumor. Tumor is hyperintense to the surrounding muscle. The tumour may be confused with glandular and lymphoid tissues of similar intensity. Comparison with T1-weighted scans overcomes this problem. Tumour, peritumoural edema and inflammation are also indistinguishable [65]. This causes T2-weighting to overestimate tumour dimensions [67-68]. Recent trauma, often due to excisional biopsy or radiotherapy, increases the amount of peritumoural edema and further obscures results.

2.6.3.3 Contrast enhancement

Post-intravenous gadolinium contrast T1-weighted images show enhancement of OCSCC, due to the increased vascularity. Contrast is best when tissue fat content is low or suppressed. Contrast allows the tumor to be well demarcated from peritumoural edema. However, there is evidence of contrast overflow into edematous tissues if imaging is delayed following administration [67]. Normal tissues, such as soft palate, mucosal, glandular and lymphoid tissues in the tonsils and tongue base also enhance and obscure adjacent tumour [63, 68]. An example is shown in Figure 2.22.

2.6.3.4 Bony invasion

Invasion of cortical bone is seen as an interruption or erosion of the border between hyperintense tissue and the signal void of cortical bone. Extensive lesions breach the cortical bone exterior and invade medullary bone. In these cases, the medullary fat is replaced by tumour tissue. On T2-weighted imaging the tumour tissue is higher signal than the medullary fat, while tumour appears hypointense on T1-weighted images. Gadolinium contrast has a detrimental effect on imaging of medullary invasion and makes invasive tumour isointense to marrow fat on T1-weighted images [64].

2.6.3.5 Imaging plane

The most useful imaging planes are axial and coronal. Both planes give information about the location and size of the tumor as well as extension across the midline [67]. The sagittal plane may be used for lesions of the oral tongue.

Figure 2.22: Coronal MRI view of oral cavity with infiltrating squamous cell carcinoma on the lateral aspect of the tongue. On the T1-weighted image (top left) normal tongue is seen on the right side as an area of high signal intensity fat content. Mass effect, due to the left-sided tumour, is seen as an area of bulging on the left dorsal aspect of the tongue. A gadolinium enhanced T1 study (top right) gives the lesion less contrast with normal tissue. A fat suppressed, gadolinium enhanced T1-weighted image (bottom right, different case) clearly shows the lesion, outlined by arrowheads. A T2-weighted study of a right sided lesion clearly demonstrates the extent of the tumour. Source: [66]

2.6.4 MRI artifacts affecting oral cavity imaging

Motion and susceptibility artefacts represent the common sources of image degradation in oral imaging. Other artefacts, while they may be present, are less common and not considered here.

2.6.4.1 Susceptibility artefact

Susceptibility artefact occurs when there is a large change in the magnetic susceptibility of the region of interest that disrupts the local magnetic field. Ferromagnetic substances cause large variations in the local magnetic field, leading to differences in precessional frequency and rapid dephasing of the free-induction decay (FID) envelope. Ferromagnetic devices, such as root canal prostheses and metal bridgework, result in large distortions of the oral anatomy [65]. Dental amalgam does not cause susceptibility artefact. This is superior to CT where dental amalgam cause beam hardening [67].

2.6.4.2 Motion artefact

Motion artefact results from voluntary and involuntary patient motion. The artefact occurs along the phase encoding gradient direction since phase samplings are separated by a lengthy TR interval. Motion artefact is ubiquitous, and results in a loss of image contrast resolution [12]. Swallowing is the primary source of motion artefact in oral MRI.

2.7 TNM staging of oral cancer

TNM staging is central to the work of this thesis. This section considers the principles of TNM staging, TNM criteria for the oral cavity, the process for change and notable shortcomings.

2.7.1 Principles of TNM staging

The Manual for Staging of Cancer, is currently in its 7th edition and was released in 2009 [69]. As stated by the American Joint Committee on Cancer (AJCC), the precise clinical and histopathological description of malignancies is central to five related objectives.

> *The selection of primary and adjuvant therapy.*
> *Estimation of prognosis.*
> *Assistance in evaluation of the results of treatment.*

Facilitation of the exchange of information among information centers.
Contribution to the continuing investigation of human cancers.

The TNM system meets these objectives while establishing a standardized language for international communication. TNM staging is primarily based on the anatomical extent of disease. This is based on the principle that:

> *The anatomic extent to which the disease has spread is probably the most important factor determining prognosis and must be given prime consideration in evaluating and comparing different therapeutic regimens.*

The ''T'', ''N'' and ''M'' refer the increase in local extent of the primary tumour, the spread to regional lymph nodes and the development of distant metastases. The T, N and M are combined with numbers to indicate increasing extent of disease. In cases where the extent of disease is unknown, the number is replaced by ''X''. In cases where *carcinoma in situ* is present in the absence of invasive disease, the suffix ''is'' is used. Possible values include:

TX, Tis, T0, T1, T2, T3, T4

NX, N0, N1, N2, N3

MX, M0, M1

Lower case letters may follow a number to indicate subdivisions within a classification set. For example, N2 $\rightarrow$ N2a, N2b.

Based on the combination of T, N and M, stage grouping is assigned. This is to account for the fact that tumours differing in anatomic extent may have a similar prognosis. For example, a T3N0M0 and a T1N1M0 may both be designated ''Stage III''. Stage grouping values range from 0 to IV.

With each new phase of cancer treatment, new information with prognostic value is made available. Lower case letters are used as prefixes to describe the four possible types of classification.

Clinical TNM	→	*cTNM or TNM*
Pathologic TNM	→	*pTNM*
Restaging TNM	→	*rTNM*
Autopsy TNM	→	*aTNM*

The cTNM is assigned prior to definitive treatment and based on clinical examination, biopsy or imaging. cTNM is used to assign a preliminary prognosis and to direct therpy.

The pTNM uses information gained during surgery and pathological examination to supplement. In many cases clinical examination can under or overestimate tumour size. Similarly, regional and distant metastasis may be subclinical and not visible on imaging. The pTNM is used to refine prognosis and direct adjuvant therapy.

The rTNM is assigned when a prior to treatment of a recurrent tumour and incorporates current plus all prior staging information. The aTNM is assigned in cases when cancer is found post-mortem.

2.7.2 Changing the TNM

Because of the nature of the TNM system as a multidisciplinary international and historical standard for cancer classification, there is a great deal of associated inertia. One of the biggest issues is that data gathered under a new classification scheme is not comparable with historical data. The AJCC acknowledges that there is ''danger'' associated with change and states that:

> *...only factors validated in multiple large studies have been incorporated into the staging system [13].*

In an age of widespread use of evidence based medicine, the revision process is often viewed as one of the TNM system's major shortcomings [70].

2.7.3 TNM criteria for the oral cavity

The current TNM definitions are set out in Table 2.5, while the associated stage groupings are shown in Table 2.6. Despite significant advances in the management of oral cancer since Denoix began his

work in 1943, there have been few major changes to the TNM criteria.

Table 2.5: TNM staging criteria for the oral cavity

Primary Tumor (T)	
Tx	*Primary tumor cannot be assessed*
T0	*No evidence of primary tumor*
Tis	*Carcinoma in situ*
T1	*Tumor less than 2 cm in greatest dimension.*
T2	*Tumor larger than 2 cm but less than 4 cm in greatest dimension*
T3	*Tumor larger than 4 cm in greatest dimension*
T4a (oral cavity)	*Tumor invades adjacent structures (cortical bone, extrinsic tongue muscles, maxillary sinus, facial skin)*
T4b	*Tumor invades masticator space, pterygoid plates or skull base or encases the internal carotid artery*
Regional Lymph Nodes (N)	
Nx	*Regional nodes cannot be assessed*
N0	*No regional lymph node metastasis*
N1	*Metastasis to a single ipsilateral lymph node, 3 cm or less in greatest dimension*
N2a	*Metastasis in a single ipsilteral lymph node larger than 3 cm but less than 6 cm in greatest dimension*
N2b	*Metastasis in multiple ipsilateral nodes, 6 cm or less in greatest dimension.*
N2c	*Metastasis in bilateral or contralateral nodes 6 cm or less in greatest dimension*
N3	*Metastasis in a lymph node larger than 6 cm in greatest dimension.*
Distant Metastasis (M)	
Mx	*Distant metastasis cannot be assessed.*
M0	*No distant metastasis.*
M1	*Distant metastasis.*

Table 2.6: Stage groupings for TNM staging of the oral cavity

	N0, M0	N1, M0	N2, M0	N3, M0	Any N, M1
T1	1	3	4a	4b	4c
T2	2	3	4a	4b	4c
T3	3	3	4a	4b	4c
T4a	4a	4a	4a	4b	4c
T4b	4b	4b	4b	4b	4c

2.7.4 The shortcomings of the current TNM staging system

The TNM staging of oral cancer has been widely reported as inadequate [71-73]. This is especially

true of the T component, which is based on the maximum superficial diameter of the lesion.

Clinical measurement of oral cancers is difficult and limited by poor access, patient discomfort and the gag reflex. Lenz et al. found that clinical appraisal of oral cancer T stage was correct only 47% of the time [74]. Diameter is also not necessarily related to prognosis. Large superficial tumors can be managed well, while some small, deep tumors evade locoregional control.

The majority of oral cancers fall in the T2 category (20–40 mm). It is unlikely that such a varied group of tumours would have similar prognosis. Tumours beyond 40 mm usually invade surrounding structures and are designated T4, leaving few T3 tumours [75].

Surgical planning depends on tumour depth. The surgeon needs to know the depth of tumour invasion and extension into surrounding oral tissues to achieve adequate margins.

2.7.5 Beyond the use of superficial diameter

MRI accurately reports T stage 87% of the time [67]. However, this does not address the fundamental issue with the T staging criteria. The largest tumour diameter, even if determined with 100% accuracy, does predict survival or occult cervical lymph node metastasis (OCLNM). In the search for alternate prognostic criteria, depth of tumour invasion has emerged as the front runner.

2.7.5.1 Depth of tumour invasion

Depth has been reported as a predictor of OCLNM and survival [44-47, 76-100]. Despite this, there is little consensus on which depth "cut-point" is significant. Recommendations for elective neck dissection and survival vary from 1.5-12 mm.

The majority of these studies rely on the “gold standard” histological measurement of depth. However, formalin treatment causes the shrinkage of specimens by up to 33%. Yuen et al. pinned specimens to foam boards during fixation, but still estimate shrinkage at 10% [100]. Others minimize the time between formalin treatment and sectioning, or hypothesize that shrinkage is minimal in the dense tumour tissue [101].

Histopathological depth can only be measured post-resection and therefore can play no role in surgical planning. For this reason, it makes sense to use diagnostic imaging to evaluate depth. The ability of high resolution MRI to evaluate tumour depth has been validated [102-104]. All three groups correlated MRI measurement with histological measurement. Tumour depth was overestimated on both T2 and T1-weighted images in all three studies.

2.7.5.2 Volume of tumour invasion

The less explored relationship between tumour volume and the prediction of survival and OCLNM follows similar principles to depth; however it is under-explored in the literature.

Yuen et al. considered volume, thickness and several other parameters [100]. Those patients with cancer of the tongue base, other oral sites, or stage T4 were not included. Samples from 85 patients who underwent surgical resection of tongue carcinoma were analyzed. Volumes were calculated using digital measurement of tumor area on histological slices, multiplied by the slice thickness and the number of slides with tumor present. No significant relationship was found

Preda et al. also looked at volume [101]. Their study included resected samples from 33 patients. Volumes were calculated by assuming an ellipsoid shape. The length of the three orthogonal tumour axes, a, b and c, were measured on axial and coronal T2-weighted studies and volume calculated using the formula $V = (4/3)\pi abc$. No significant relationship was found.

Chew et al., however, had more success [105]. The group found that volume, using a cut-off value of 13.0 cm^3 was predictive of ACS and DFS. However, the sample size used was only 17 cases of lingual SCC.

These methods are not without their limitations. The methods used by Yuen et al. are subject to the usual shrinkage of samples in formalin, despite steps to avoid this [100]. Preda et al, while using MR images, assume an ellipsoid shape to calculate volume, while the truth is that tumours assume a variety of non-standard shapes and growth patterns [101].

2.8 Prognostic and predictive factors

Tumour and patient-related factors for the prediction of survival and OCLNM are also an important part of this work. A background for each factor is essential to understanding univariate analysis in Chapter 5, and model building in Chapter 6.

2.8.1 Patient factors

2.8.1.1 Age

The mean age at diagnosis for OCSCC is approximately 60 years [106]. Several authors have reported a rising incidence of oral cancer in young adults, especially the oral tongue [107-110]. Prognosis seems to be worse in younger patients [111]. However, more recent papers that adjust for confounding variables find that prognosis is no worse than that of older cohorts [110, 112].

Many authors conclude that there is no relationship between age and survival [46, 76, 113-114]. Conversely, others conclude that prognosis worsens with increasing age [15, 115-116]. Data from the UK Office for National Statistics, shown in Figure 2.16, shows the effect of increasing age on survival. There is no evidence of a relationship between age and the presence of CLNM [81, 98, 117].

2.8.1.2 Gender

The gender distribution of oral cavity cancer is predominantly male with a ratio of 2:1 [106]. This underscores a trend that has seen the ratio decrease from approximately 5:1 in the 1950s. In the period from the mid-1980s to 1999, the incidence of OCSCC in men decreased while remaining steady in females [118]. Many authors point to an increase in smoking and alcohol consumption among females [119]. Cancer Research UK statistical figures show that there has been a slight increase in the male-to-female ratio of oral cavity carcinoma incidence in the period from 1975 to 2005 [120].

A review of the literature by Massano et al. [121] found several studies suggesting that gender was not associated with survival outcomes [46, 113-114]. In a study by Leite and Koifman, an apparent decrease in survival among females was attributed to delays in seeking medical care and increased

refusal of definitive treatment [115]. There is no evidence of a relationship between patient gender and the presence of CLNM [81, 98, 117].

2.8.1.3 Alcohol and Smoking

Alcohol consumption and smoking are well known OCSCC risk factors. In a recent study of 264 patients by Mayne et al., 111 patients (42.0%) were former smokers and 136 patients (51.5%) were current smokers [122]. Only 17 (6.4%) were never smokers. Similarly, 125 of the 264 patients (47.3%) consumed in excess of 21 units of alcohol per week.

Pre and post-diagnosis smoking and alcohol consumption have been shown to decrease overall survival [122-123]. In a recent study by Mayne et al. there was a dose-dependent response between the number of pack-years smoked and all-cause mortality (RR 5.39 for >60 pack-years). A dose-dependent response also existed for alcohol consumption (RR 2.36 for >21 drinks/week).

Current smoking at the time of diagnosis and continued alcohol consumption post-diagnosis have been found to increase the risk of second primary tumours [124]. Mansour et al. noted a correlation between smoking and increased risk of CLNM [125]. Though nothing could be found studying alcohol consumption directly, Mansour noted that alcohol consumption and smoking were highly correlated in their cohort [125].

2.8.1.4 Adult Comorbidity Evaluation - 27 (ACE-27)

Several authors have already demonstrated that comorbidity is a significant predictor of outcome in head and neck cancer [126-128]. The Adult Comorbidity Evaluation - 27 (ACE-27) system was developed by Piccirillo [129]. It is a derivative of the Kaplan-Feinstein Index (KFI), originally developed to assess comorbidity in diabetes mellitus patients [130]. The ACE-27 uses a series of criteria, based on information commonly found in patient records, to classify comorbid illness into grade 1 (mild decompensation), grade 2 (moderate decompensation) or grade 3 (severe decompensation) disease. The categories assessed by the index include:

Cardiovascular system	Psychiatric
Respiratory system	Rheumatologic
Gastrointestinal system	Immunological system
Renal system	Malignancy
Endocrine system	Substance abuse
Neurological system	Obesity

Based on the grade assigned to each comorbid illness, an overall comorbidity score of mild, moderate or severe is assigned. The score is assigned according to the highest ranked illness, except where two or more grade 2 illnesses occur in different categories, in which case the overall score is designated "severe". The entire ACE-27 comorbidity index is found in Appendix I.

The ACE-27 was developed to be used for retrospective chart review and has been shown, in the US, to be a useful tool for retrospective studies [131]. Paleri et. al found that information commonly found in head and neck cancer records in the UK was sufficient to apply the ACE-27 [132].

There are no reports of a relationship between comorbidity and CLNM. There are no reasons for patients with comorbid illness to have more aggressive tumour phenotypes, unless their illness is associated with immune suppression or predisposition to malignancy.

2.8.2 Tumour factors: clinical

2.8.2.1 Oral cavity subsite

The oral tongue, floor of the mouth and mucosal lip are the most frequently involved subsites and account for 70-80% of all OCSCC. The frequency of involvement for each subsite is given in Figure 2.23 and Figure 2.24.

The distribution of vascular and lymphatic vessels varies between oral cavity subsites and is thought to influence the incidence of CLNM and survival [121]. Some oral cavity subsites are also more challenging to access surgically, such as the superior gingivolabial sulcus [133].

While the theoretical basis for varying rates of metastasis and survival among oral cavity sites appears sound, few studies exist to support this conclusion. This is likely due to the large number of

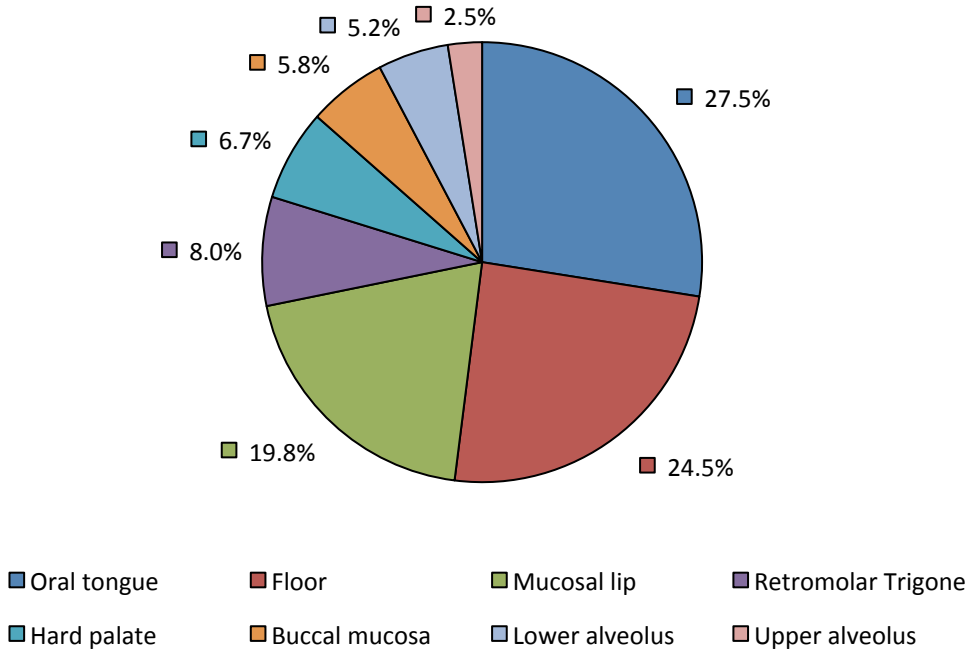


Figure 2.23: Tumour frequency by site: National Cancer Data Base (NCDB). Source: [134]

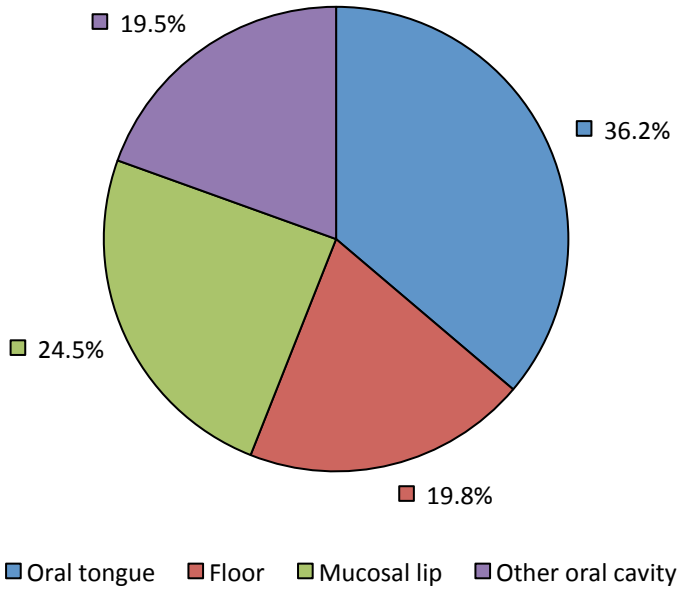


Figure 2.24: Tumour frequency by site: Surveillance, Epidemiology and End Results Program (SEER). Source: [134]

oral cavity sites, as well as the relative rarity of disease at sites other than the oral tongue, floor and lip. Leite showed higher mortality in patients with tongue carcinoma compared to lip carcinoma [115]. Higher rates of metastasis have been reported for the tongue base relative to the anterior two-thirds [133].

2.8.2.2 Growth pattern

The gross appearance of OCSCC can be grouped into four categories; exophytic, endophytic, ulcerative or superficial spreading. Ulcerative lesions are characterized by an irregular edge and induration. Exophytic growths can appear as cauliflower-like projections or white, heaped up lesions with varying levels of surface keratin. Endophytic tumours have small surface components but a large degree of underlying tissue involvement. Superficial proliferative lesions have large surface components with little underlying soft tissue involvement [3].

Exophytic tumours are known to have improved locoregional control rates, as well as improved survival, relative to endophytic and ulcerative tumours [3]. Investigators agree that exophytic tumours behave like thin tumours and depth measurements should be taken from reconstructed mucosal surfaces, ignoring the exophytic component [135]. Fukano et al. concluded that gross tumour morphology did not predict CLNM in a sample of oral tongue carcinomas [81].

2.8.3 Tumour factors: staging

2.8.3.1 T-classification

The primary tumour is clinically assessed at the time of presentation. During the examination, maximum surface diameter is measured using a ruler and used to assign the clinical T (cT). The lesion is also palpated to gauge the depth and invasion into critical structures like extrinsic muscles.

Clinical examination of the primary tumour is supplemented by imaging. Based on maximum diameter measurements, the cT value may be revised. Pathological measurements of the diameter are taken post-surgically on the resected specimen.

The criticisms of diameter-based T-staging have been discussed in Section 2.7.4.

2.8.3.2 N-classification

Cervical lymph node metastasis is assessed clinically by careful palpation at presentation. Imaging has been shown to be more sensitive for nodal enlargement [136-138]. A variety of imaging techniques are used to evaluate CLNM, including MRI, CT, ultrasound and FDG-PET. This work focuses on MRI and the reader is directed to Simon and Rubinstein for alternative techniques [64].

On MRI, metastatic lymph nodes are distinguished from reactive nodes using a variety of criteria. Size is the primary criterion, with a lymph node being suspicious of metastasis if the maximum axial diameter exceeds 10 mm [139]. Other size criteria have been suggested [140]. Using 10 mm, MRI has been reported to have a sensitivity of 82% and specificity of 48% [141]. The presence of necrosis increases the probability of metastasis and appears as intranodal regions of high signal on T2-weighted scans [139, 142] and low signal [139] or no contrast enhancement [142] on T1-weighted scans. The loss of sharp borders or invasion into surrounding tissue is suggestive of extracapsular spread (ECS). Both necrosis and ECS are relatively specific markers but lack sensitivity [121].

Other indicators that raise the suspicion of metastasis are the presence of groups of enlarged nodes in the expected drainage distribution of the primary tumour. A rounded shape, loss of hilar fat or indentation, or a ratio of long to short axis of less than 2 have been suggested as criteria [143].

CLNM is associated with a 50% decrease in survival and increased recurrence rate. In cases of OCSCC with no clinical or imaging evidence of metastases, CLNM was detected in 20-40% of pathological specimens [50].

2.8.3.3 M-classification

Distant metastases are rare in oral cancer. A Stanford study of 217 patients reported that 17% of patients developed distant metastases, while a large data set from MD Anderson Cancer Center of 5019 patients reported a value of 11% [144-145]. Both studies had a majority of advanced cases.

Common sites included bone, brain, liver and lung. Distant metastases are investigated by chest radiograph or chest CT. PET and PET-CT are not in routine use for staging.

2.8.3.4 Stage

Many authors are critical the current TNM criteria. However, the majority accept that there is a role for TNM staging. Guerra et al. reported that the 5-year survival of Stage I and II SCCs was 82% versus 49% for Stage III and IV disease. Lo et al. reported 5-year survivals of 75%, 66%, 49% and 30% for Stage I, II, III and IV respectively [113].

2.8.3.5 Extracapsular spread (ECS)

Extracapsular spread (ECS) refers to the extension of metastatic disease beyond the confines of the lymph node capsule [121]. ECS can be subdivided into macroscopic, visible to the naked eye, and microscopic, seen only during histological analysis [146]. An example is shown in Figure 2.25.

ECS is a predictor of distant metastasis, locoregional recurrence and survival [50, 146-147]. Woolgar et al. have shown that the three-year survival of 173 OCSCC patients with nodal metastasis was 72% for those without ECS versus 33% for those with macroscopic ECS [147]. Woolgar also showed that there was no difference in survival between microscopic and macroscopic ECS [147].

2.8.3.6 Bone invasion

The association between bone invasion and prognosis is poorly defined; however its inclusion in the OCSCC TNM criteria implies that it does have some impact [148]. The consensus among those in charge of developing the original OCSCC TNM criteria was that bone invasion was associated with a poor clinical course [106]. Some authors have confirmed this opinion, showing a relationship between bone invasion and survival [149]. However, others have disputed the prognostic value of bone invasion [149].

Imaging is superior to clinical examination, which has a reported sensitivity and specificity of 61% and 81% respectively [150]. In a meta-analysis by Brown et al, CT had a mean sensitivity and specificity of

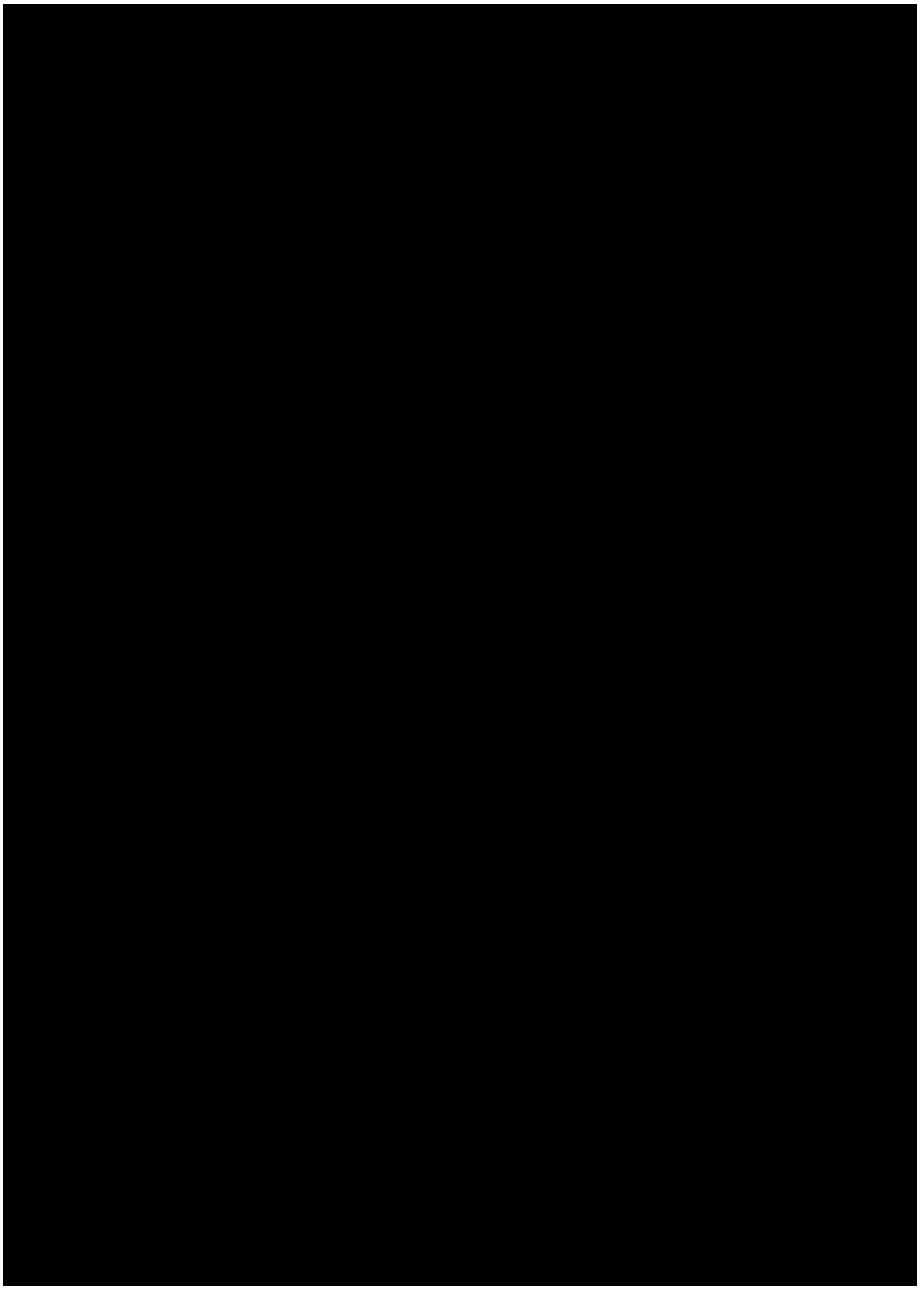


Figure 2.25: Extracapsular spread. In the upper image, metastatic tumour is contained within the lymph node. Uninvolved perinodal fat is marked with an asterisk. In the lower image, microscopic extension beyond the lymph node capsule has occurred. Thin arrows indicate remnants of capsule. Source: [121]

86% and 75% respectively, compared to values of 72% and 85% for MRI [150]. Bone invasion is definitively assessed by pathological analysis.

2.8.3.7 Midline involvement

Tumours that cross the midline are likely to be advanced. Further, patients with tumours that cross the midline are at increased risk of contralateral CLNM. Kowalski et al. showed that tumours of the floor, as well as those within 1 cm of the midline, are at increased risk of contralateral CLNM [151].

In a series of 53 oral tongue SCC patients under the age of 35, Mallet et al. found that midline invasion was an indicator of decreased overall survival but not disease-free survival [152]. Similarly, Weber et al. studied 188 patients with SCC of the oropharynx, concluding that survival was poorer among patients with midline invasion [153].

2.8.4 Tumour factors: histopathology

2.8.4.1 Tumour differentiation

The tumour grade is assigned based on the similarity of the tumour to normal epithelium. Grade description follows the conventions outlined in the World Health Organization (WHO) classification [154]. The most aggressive area of tumour is viewed at medium magnification and is described as either well, moderately or poorly differentiated. Grade has been discussed previously in Section 2.4.4.

This grading system is widespread throughout oncology but its subjective nature leads to problems with interobserver variability [154]. Most studies demonstrate that lower degrees of tumour differentiation are associated with decreased survival [113-114, 155-157]. However, some authors refute this claim [46, 76, 115]. There is also association between a lower degree of tumour differentiation and increased risk of CLNM [79, 94, 98, 158-159]. This trend was not noted by others [47, 81, 117].

2.8.4.2 Infiltrative front

Authors describe two distinct patterns of infiltration. Cohesive tumour fronts demonstrate pushing, well defined tumour borders [106]. Non-cohesive tumour fronts demonstrate finger-like projections, cell islands or single cells at the periphery. In literature reports, a non-cohesive tumour front was found in the majority of cases with frequency ranging from 79-95% [39-40, 160]. Examples of each type of invasive tumour front are shown in Figure 2.26 and Figure 2.27.

Several authors have noted an increased rate of CLNM when non-cohesive fronts are present [39-40, 160]. Similarly, authors have noted that non-cohesive fronts are associated with decreased survival [39]. Other authors have found no such association [161-162].

2.8.4.3 Lymphovascular invasion (LVI)

Lymphovascular invasion is defined as the presence of malignant cells within an endothelial lined space. Lymphatic channels need not be differentiated from venous vessels. In the literature, the reported incidence of LVI varies widely from 15-51% [87, 117, 163]. An example is shown in Figure 2.28.

Several authors have reported that LVI is a predictor of CLNM [39, 78, 94, 98, 117, 164-166]. Close et al. believe that LVI was the most important predictor of CLNM [161]. Crissman et al. found LVI to be the only predictor of CLNM on multivariate analysis [39]. However, the work of other authors refutes these claims [47, 160, 162, 167-169].

There are fewer studies that look at recurrence and survival. However, Yuen et al. and Fagan et al. noted no association between with local recurrence [47, 170]. Rahima et al. noted no association with local, regional or distant recurrence [163]. Clark et al. reported no association with disease-specific survival [169].

2.8.4.4 Perineural invasion (PNI)

Perineural invasion refers to the invasion of the perineural space ahead of the invasive front of the

carcinoma [154]. PNI within the tumour is of little prognostic significance and not reported [154]. The prognostic significance of PNI likely stems from the role of the perineural space as a route of tumour extension [171]. Literature reports indicate that the incidence of PNI ranges from 6-30% [78, 172-174]. PNI is of particular importance in another tumour of the oral cavity, adenoid cystic carcinoma. An example is shown in Figure 2.29.

PNI is closely associated with CLNM in OCSCC [98, 163, 170, 174-175]. Numerous studies have also found a relationship with local recurrence [163, 170, 173, 176-179], regional recurrence [78, 84, 98, 163, 173] and survival [47, 78, 163, 170, 172-174, 176, 178, 180]. Studies that show a correlation between PNI and tumour grade hint that PNI is associated with a more aggressive tumour phenotype.

2.8.4.5 Surgical margins

The distance from neoplastic tissue to the deep and surgical margins are measured histologically. Based on the distance obtained, the margins are classified as clear if the margin is greater than 5 mm, close if the margin is from 1-5 mm and involved if the margin is less than 1 mm.

The presence of involved margins is associated with increased incidence of local and regional recurrence, as well as reduced disease-related survival in OCSCC [181-185].

2.9 Conclusions

The purpose of this chapter was to arm the reader with a working knowledge of the concepts required to interpret the work found in the remainder of this thesis. Though this background assumes a basic level of expertise in oral oncology, an attempt has been made to direct the interested reader to sources that provide information beyond the scope of this text.

This summary of oral cavity anatomy, oral cancer, MRI imaging and imaging of the oral cavity is a useful reference while reading the remainder of the text and, when helpful, the text will internally reference sections of the background.

Figure 2.26: Cohesive infiltrative front. Tumour growth is well defined with pushing borders. Source: [186]

Figure 2.27: Non-cohesive infiltrative front. This chaotic pattern demonstrates finger like projections, cell islands or single cells. Source: [186]

Figure 2.28: Lymphovascular invasion. Tumour cells with large, darkly staining nuclei are seen invading a blood vessel. Source: [177]

Figure 2.29: Perineural invasion. In "A", there is "onion skin" layering of tumour cells around the nerve. "B" shows a crescent of tumour cells superior to the nerve. "C" shows tumour cells encircling a longitudinal nerve section. "D" shows tumour cells contacting the nerve on two surfaces, left and right. "E" shows tumour cells surrounding the nerve. "F" shows intraneural invasion. Source: [170]

3 General methodology

3.1 Introduction

In the previous chapter the relevant background was discussed in detail. This next chapter deals primarily with methodology. In some cases, the methodology is best left until the relevant chapter of the text. However there is a body of methodology that is common to all aspects of the thesis and best outlined before proceeding further. Like the previous chapter, this chapter should serve as a reference for the reader as the remainder of the text is explored.

3.1.1 Chapter goals

The goals of this chapter are:

1. *To provide the reader with a summary of computer hardware and software used throughout the thesis.*

2. *To outline the procedure for database construction and management, including patient identification, data security, MRI study retrieval and associated problems.*

3. *To summarize the cases excluded from the present study as well as to summarize the cases that comprise the final study sample.*

4. *To outline the procedure for the measurement of MRI-based volume, thickness and diameter.*

5. *To outline the procedure for collection of data for each prognostic and predictive factor included in this work.*

3.1.2 Organization of sections

The chapter begins with a brief outline of computer hardware and software used in this work. The purpose and relevance of each software is given.

The next three sections are focused on database construction and management. This constitutes a major part of this work and is the foundation upon which the remainder of the thesis rests. Important topics include the definition of inclusion and exclusion criteria, candidate identification and information retrieval, database design and data input, MRI study retrieval and the implementation of data security procedures. While perhaps suitable for discussion in future

chapters, the problems and dilemmas associated with database construction are covered here, allowing the reader to contemplate the implications in context. A summary of included and excluded cases is given.

The final two sections detail methodology specific to each prognostic and predictive factor that was included in the database. Because the focus of this thesis is on the contribution of MRI-based tumour morphometry, the relevant procedures are given a section of their own. Here, general methodology for image processing and measurement are outlined. The specifics of volume, thickness and maximum diameter measurement are then detailed. This is followed by the remainder of the prognostic and predictive factors. Their primary information source within the patient record is given as well as relevant clinical or histopathological procedures.

3.2 Hardware and software

This section details hardware and software that were central to the work carried out in this thesis.

3.2.1 Hardware and operating systems

All image viewing and analysis was done on a workstation equipped with an Intel Celeron 2.80 GHz CPU, 1.00 GB of RAM and a Radeon 9550 graphics card (ATI Technologies Inc.). Each image was viewed on a 19" liquid crystal display (LCD) monitor (CTX International) at a resolution of 1280 X 1024 pixels and 32 bits of color depth. The monitor was not calibrated for brightness and contrast prior to use. The operating system was Microsoft Windows XP Professional (Service Pack 3).

3.2.2 Database management software

TrueCrypt (Version 6.1, TrueCrypt Foundation) is an open source software that allows the user to create encrypted volumes. Encrypted volumes were required on all USB memory drives used to transfer patient data.

Cobian (Amanita, Version 9.5.0.201, Luis Cobian) is an open source data backup software. Backups can be encrypted using sophisticated security algorithms.

Eraser (V 5.86, The Eraser Project) is an open source software that safely and permanently erases sensitive data from a hard drive.

Microsoft Office Access and Excel 2007 (Microsoft Inc., Redmond, USA) are commonly used database and spreadsheet software.

3.2.3 Imaging and measurement software

ImageJ (Version 1.41o, National Institutes of Health, USA) is a Java-based public domain software that is used for image processing research. ImageJ was used for all MRI-based measurements.

eFilm (Version 2.1.0, Cedara Software) is a proprietary DICOM viewing and PACS server software. eFilm was used to maintain an MRI archive and to assess image quality prior to measurement.

K-PACS is a free proprietary DICOM and PACS server software. It's function is similar to eFilm, but it includes a DICOM anonymization function. This function was used for all MRI series anonymization.

3.2.4 Statistical software

R-project (V. 2.9.1, The R Foundation for Statistical Computing) is an open source statistical software that is widely used in the research community. The R programming language allows the user to create custom "scripts" that are useful for data analysis and simulation. User submitted packages extend the functionality of R and are available through the Comprehensive R Archive Network (CRAN). R was used for all statistical simulation, statistical modeling and some graphical output.

The Statistical Package for Social Sciences (SPSS, V. 16.0.2, IBM Corp., New York, USA) is a proprietary software package commonly used in the medical sciences. SPSS was used for all univariate analysis.

3.3 Inclusion and exclusion criteria

Patient information was collected from two tertiary care centers: The John Radcliffe Hospital (JRH), Oxford and University College London Hospital (UCLH). Patients diagnosed with histologically proven OCSCC presenting 1998 to present were studied retrospectively. All patients undergoing staging MRI

in whom the primary treatment was surgical resection with curative intent were included.

No patient underwent pre-operative chemotherapy or radiotherapy. Patients were excluded if their MRI studies could not be recovered or were incomplete, if they underwent tumour biopsy within seven days prior to MRI (to avoid false positive findings from inflammatory response), or if tumours were obscured by motion or susceptibility artefact. Patients were also excluded if they had previous OCSCCs, multiple synchronous OCSCCs, or missing or incomplete records. Patients with primary SCC of the lip were excluded as these lesions are rarely imaged. Inclusion and exclusion criteria are summarized in Table 3.1.

Table 3.1: Inclusion and exclusion criteria

Inclusion Criteria
Presenting to either institution from 1998 to present
Histologically proven primary squamous cell carcinoma of the oral cavity
Primary treatment surgery with curative intent
Staging MRI study
Exclusion Criteria
Pre-operative chemo- or radiotherapy
MRI studies not recovered or incomplete
Underwent tumour biopsy within 7 days prior to MRI
MRI studies of poor quality due to motion or susceptibility artefact
Previous or synchronous OCSCCs
SCC of the lip
Missing or incomplete patient records.

3.4 Database construction and management

In this section, database construction and management is outlined including patient identification, data gathering, MRI recovery and data security. Associated problems are discussed.

3.4.1 Patient identification and record retrieval

3.4.1.1 John Radcliffe Hospital, Oxford

Patients undergoing OCSCC resection were identified from two primary sources. A cancer registry maintained by the Department of Oral and Maxillofacial Surgery was available from 1998 to 2006.

This registry detailed patient identifying information, as well as the histological type and location of the cancer. For 2006 to present, Departmental theatre records were used to identify patients. These records detail identifying patient information as well as the nature of the surgery.

The patient identifying information was confirmed using the patient archiving system (PAS). In some cases, some aspect of the patient identifying information was missing. In cases involving patient name, the PAS was used to find the correct information. In cases involving the hospital identification number, the patient name and/or date of birth was used to search for the correct patient on the clinical intranet. Patients were identified as ''correct'' based on matching first and last names and a record of a prior ''episode'' involving Oral Surgery.

Once a list of potential patients was compiled, paper records were screened for exclusion and inclusion criteria. Until recently, patient files were grouped according to the department providing care. Department of Oral and Maxillofacial Surgery records were stored in a separate archive and designated ''Oral''. Oral cancer patients often receive multidisciplinary care and, as a result, had multiple paper records from Oral and Maxillofacial Surgery, ENT Surgery, Plastic Surgery and Radiation Oncology. In addition, patients may have ''General'' notes covering other services.

From 2008 onwards, all information for new patients is recorded in ''general'' notes. Existing ''oral'' notes are being transferred as patients are seen in clinic or theater. The records of deceased patients are stored off-site at Upper Heyford and available using next-day delivery.

3.4.1.2 University College London Hospital

At UCLH, cases were identified through review of electronic multidisciplinary oncology team (MDT) notes available through the clinical intranet. This database was available from 2006 and contained a mixture of patients with tumours of different histological subtypes, head and neck sites and treatment plans. The validity of the patient identifying information was confirmed using the clinical intranet. Patients with missing or incomplete identifying information were followed up using a procedure similar to that used at the JRH.

Several hundred MDT notes were screened for inclusion and exclusion criteria. In cases where exclusion and inclusion criteria were unclear, the case was retained for further screening. Records prior to 2006 were not immediately available. Thus, the UCLH sample spans 2006 to present. For cases identified for further screening, electronic records were recovered from the clinical intranet. Where electronic notes were unavailable or incomplete, the paper records were requested from external storage. At both institutions, electronic medical records were available from 2006 onwards.

3.4.2 Data gathering and input

Those records meeting inclusion and exclusion criteria were audited and relevant information recorded in a comprehensive database. Patient information was gathered from several categories, listed in Table 3.2.

Table 3.2: Information gathered from case review.

Patient Identifying Information	Pathologic Information
First and last name	Depth of invasion of primary tumour
Hospital identification number	Maximum diameter of primary tumour
Date of birth	Presence of bony, perineural and perivascular invasion
Sex	Tumour grade and infiltrative border
	Number and levels of cervical lymph nodes dissected
Clinical Information	Number, level and maximum diameter of metastatic nodes
Smoking status	Number and level of lymph nodes with extracapsular spread
Alcohol consumption	Pathologic TNM stage
Comorbidities	
Site of primary tumour	Follow-Up Information
Clinical TNM staging	Date and location of recurrence
Biopsy date	Date and location of second primary
	Date and cause of death
Imaging Information	Date of last known follow-up
Date of staging MRI	
Radiologist report of staging MRI	Surgical Information
Date of staging CT (if applicable)	Surgery date
Radiologist report of staging CT	Surgical approach to the primary tumour
	Surgical approach to the mandible
	Surgical approach to the cervical lymph nodes
	Adjuvant chemotherapy

Two databases were constructed. One database was reserved for patient identifying information. A

second database, the main database, was used to record information from the remaining categories listed in Table 3.2. Each case was assigned a unique identification number. This identification number was used to link the main database to the patient identifying information. This is in accordance with the principle of pseudo-anonymization, discussed in Section 3.4.4.3.

To facilitate data entry, a multi-paged form was constructed. Each form page was used for entry of each category listed in Table 3.2. A screenshot of the form is shown in Figure 3.1.

3.4.3 MRI study retrieval

3.4.3.1 John Radcliffe Hospital

MRI studies from 2006 onwards were stored on the Oxford MRI (OMRI) server. In some cases, studies were available on the JRH PACS but not the OMRI server. These files were transferred to the OMRI server.

MRI studies done from 1998 to 2006 were stored on 4 GB optical disks. These files were subsequently transferred to the OMRI server. Studies for individual patients could not be reviewed without first recovering the contents of the entire disk, a process taking approximately 20 minutes. In total, the contents of more than 200 optical disks were transferred. Once recovered, studies for individual patients were identified by searching the OMRI database using identifying information.

Patient studies were exported as DICOM files and transferred to a local folder on the main research computer. Each DICOM series was anonymized and patient identifying information replaced by a unique identification number.

3.4.3.2 University College London Hospital

Staging MRI scans were available from the UCLH patient archiving and communication system (PACS) from 2006 onwards. Studies were exported as anonymized DICOM files and transferred to a local folder on the main research computer.

3.4.4 Data security

Steps were taken to ensure the security of the information collected in accordance with the Data Protection Act (DPA) of 1998.

3.4.4.1 Data Protection Act of 1998

In 1998 the DPA was invoked as an Act of Parliament [187]. It provides legal principles governing the handling of personal information in the United Kingdom. The DPA principles are given in Table 3.3.

Table 3.3: Data Protection Act of 1998 [187]

1. *Personal data shall be processed fairly and lawfully and, in particular, shall not be processed unless:*
    a) *at least one of the conditions in Schedule 2 is met, and*
    b) *in the case of sensitive personal data, at least one of the conditions in Schedule 3 is also met*
2. *Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes*
3. *Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed*
4. *Personal data shall be accurate and, where necessary, kept up to date*
5. *Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes*
6. *Personal data shall be processed in accordance with the rights of data subjects under this Act.*
7. *Appropriate technical and organizational measures shall be taken against unauthorized or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data*
8. *Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.*

The DPA has been criticized for its complexity. The Department of Health offers this guidance concerning the use of patient information for research purposes [188]:

> *Data legitimately processed for research or statistical purposes, as long as such processing neither causes substantial harm or distress to the data subject nor is used to support measures or decisions in relation to individuals, are exempt from certain provisions of the Act. Such data can be kept indefinitely and are exempt from the subject access rights if the results of the work are not made available in a form from which data subjects can be identified. Use of such data for research, although obtained for other purposes will not breach the second*

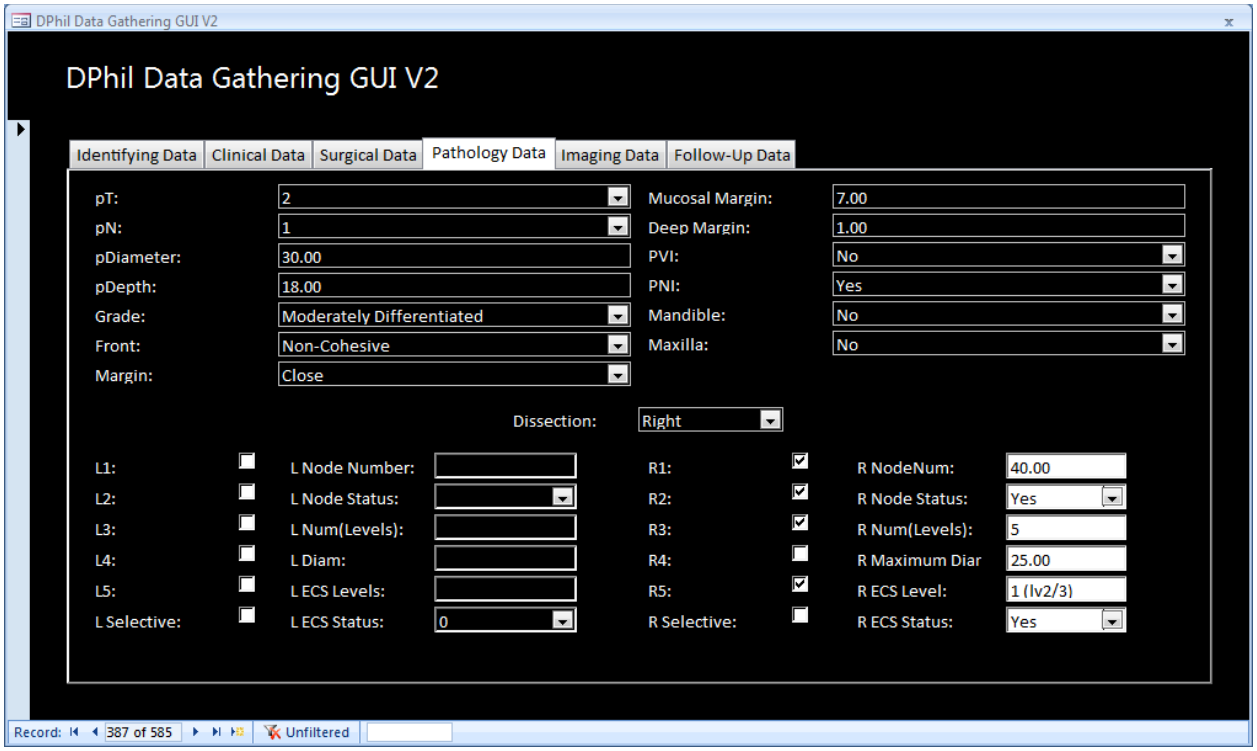


Figure 3.1: Microsoft Access data entry graphical user interface (GUI). This screenshot shows the "pathology data" component of the form for data entry. The other data categories are available as a series of tabs seen on the upper border. The form could also be used to search the database, allowing for quick and efficient review of information.

principle (use incompatible with the purposes for which it was obtained) and hence will not be unlawful on those grounds. However, this does not absolve the data controller from the obligation, in order to comply with the first principle, to give the data subject general information about intended uses.

The use of patient information in this work complies with this guidance.

3.4.4.2 Data storage and transportation

All data was stored in a secured office at the JRH. All documents containing sensitive information were stored in a locked filing cabinet when not in use. All electronic data was stored on a password-protected research computer. The computer was connected to a firewalled local area network (LAN) with broadband internet access. All databases and spreadsheets were encrypted and password protected.

When it was necessary to transfer electronic data, a 4 GB USB memory device (PNY Technologies Mini Attaché) was used. This USB drive was encrypted using TrueCrypt. A back-up of all electronic data was maintained using an external hard drive (ReadyNAS, Netgear, San Jose, USA). Cobian was used to maintain this archive. Daily incremental and weekly full back-ups were employed. Back-ups were encrypted using the 128 bit "Blowfish" algorithm.

3.4.4.3 Patient identifying information and pseudo-anonymization

The study required access to identifying information to identify cases meeting exclusion and inclusion criteria, recover patient records and recover MRI studies. This identifying data was also used to update follow-up information when necessary.

Data analysis does not require access to identifying information. ''Pseudo-anonymization'' refers to a process by which raw data is kept in a separate database from identifying information. The two databases are linked by a unique identification number and identifying information accessed only when necessary. The data required for analysis is not completely anonymous because the link to the identifying information is retained. Data security is enhanced because details of the patient's health history are revealed only when the two databases are viewed together. Loss of either in isolation is

problematic, but not catastrophic.

3.4.4.4 Data destruction

All data deletion was done using Eraser software. The database will be retained for 6 months following the end of the study period, defined as successful thesis examination.

3.4.5 Problems associated with data collection

3.4.5.1 Paper-based patient records

The quality of information varied significantly from record to record. In many cases, hand-written progress notes were illegible and alternate sources of information required. It was common for elements of the patient record to be out of order or misfiled in the incorrect section.

Patient records were often incomplete. In some cases, a small but relevant piece of information was omitted. In other cases, large portions or entire sections of information were missing, for example the histopathology report.

3.4.5.2 Decentralized record keeping

As mentioned previously, until recently the JRH used a decentralized record keeping system. As a result, patient information could be found in any of a number of possible archives. In some cases, information had to be pieced together from several sources. One result of the decentralized record keeping system was that records often became absorbed into other record archives.

Clinical records were sometimes missing entirely or may have been lost or misplaced by personnel. There may also have been a failure of PAS procedures, resulting in tracking to the wrong location. In cases where there were multiple volumes, a common problem was the loss of a single volume.

3.4.5.3 Optical disk storage of imaging studies

Recovery of MRI from optical disk is inefficient. The recovery of 200 disks at a rate of 20 minutes per disk required more than seventy hours. This is in addition to the time-cost of recovering patient studies from the OMRI server to the local host.

3.4.5.4 Labour intensive work-flow

The work flow for recovery of patient information and MRI studies, diagrammed in Figure 3.2, was time and labour intensive. For a case with all relevant information stored in a single archive and a MRI study recovered from optical disk, a conservative estimate is 30-60 minutes. Further, the first steps of the work flow (identification of potential patients) were inefficient, and several patients were identified for every one included. Considering the attrition rate, the time-cost per patient quickly increases. Anticipated time-cost per unit of the final sample is important for study design.

3.5 Summary of included and excluded cases

The exclusion and inclusion criteria were applied to many cases from both institutions. In this section, the process of compiling the final study sample is summarized.

3.5.1 Institutional distribution

A total of 541 individuals were identified using the JRH oral cancer registry and theatre records as possible SCC patients. Similarly, review of UCLH records identified 43 individuals, giving a total of 584 possible OCSCC cases.

3.5.2 Record keeping-related exclusions

Of these 584 cases, 385 cases were excluded from analysis. In 50 (13%) cases the patient record was unrecoverable. The reasons were varied. Is many cases, identifying data was recorded incorrectly or incompletely in the local cancer registry and the patient was untraceable on PAS or the clinical intranet. Alternatively, patients were identified but their records were misplaced. Of recoverable records, 8 (2.1%) were missing essential information, for example post-surgical pathology. 1 patient (0.3%) was recorded as a duplicate instance.

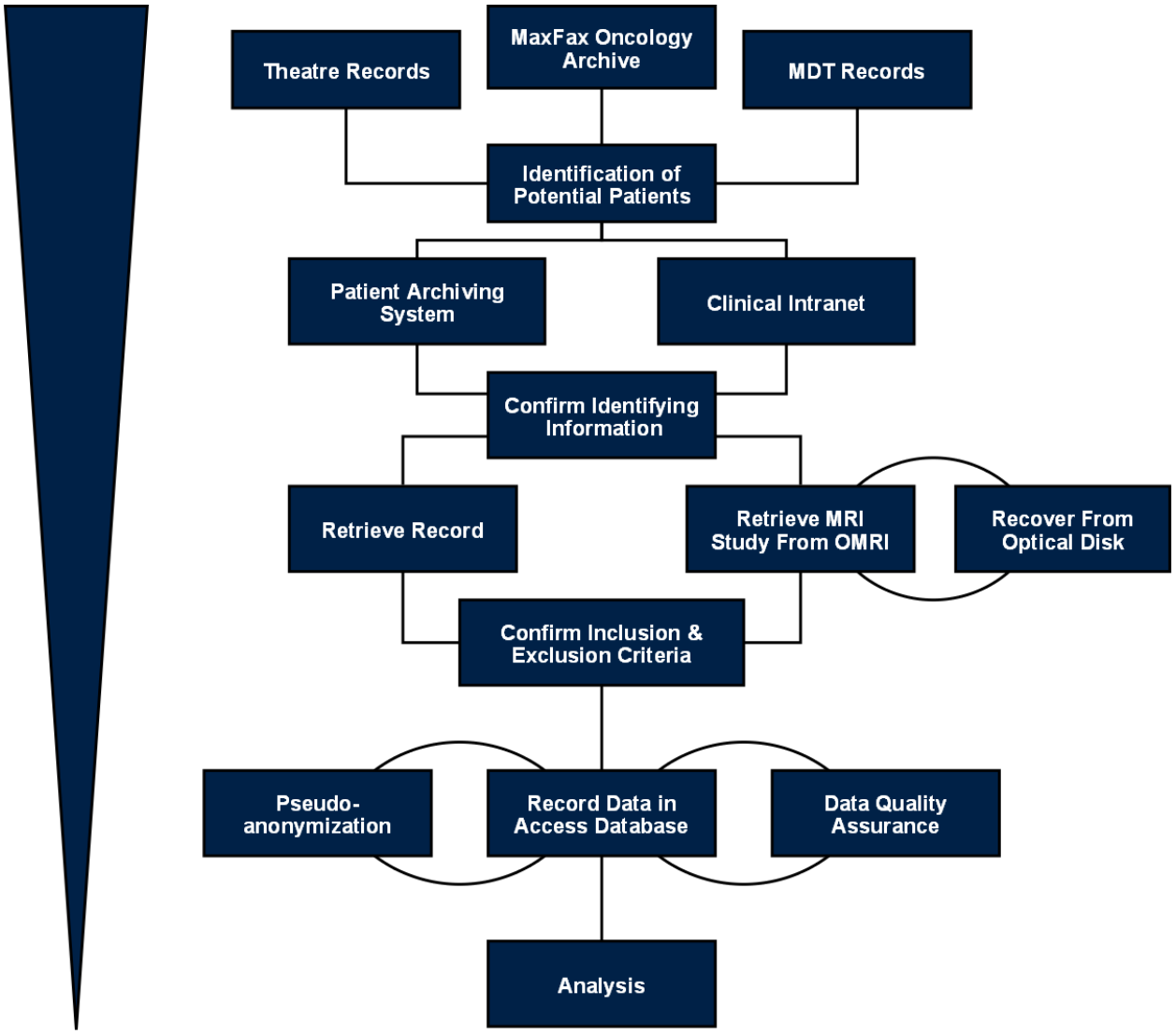


Figure 3.2: The process of data gathering. Patients initially identified as potential candidates are screened first to confirm identifying information then to confirm inclusion and exclusion criteria. Only a small percentage of cases identified as potential candidates made it to the data base entry phase. The attrition rate is represented by the inverse pyramid on the left side of the image.

3.5.3 Patients not meeting temporal and treatment requirements

25 (6.5%) patients presented with SCC prior to 1998 and were excluded. 7 (1.8%) patients received treatment before being referred to the JRH or UCLH. 5 (1.3%) patients were reviewed at the JRH or UCLH and referred elsewhere. 39 (10.1%) cases were unsuitable for primary surgical resection and referred to radiotherapy, medical oncology or palliative care.

3.5.4 Patients not meeting site and histopathological diagnosis requirements

45 (11.7%) patients were excluded because their tumours were not SCC. 37 (9.6%) were excluded because their SCC was not located in the oral cavity. 16 patients (4.2%) presented with lesions of the oral cavity that were pre-malignant. 31 cases (8.1%) presented as SCC of the lip. Also excluded were 9 cases (2.3%) with unknown primaries and 5 cases (1.3%) with multiple synchronous tumours.

3.5.5 Patients not meeting imaging requirements

73 cases (19.0%) had no staging MRI or a study that was unrecoverable. This was a problem for pre-2006 studies stored exclusively on optical disk. Of the recovered studies, 15 (3.9%) were excluded due to susceptibility and 2 (0.5%) due to motion artefact. 17 patients (4.4%) had staging MRIs done within a 7 day post-biopsy window and were excluded to avoid problems with inflammation.

3.5.6 Summary of the study sample

Of the original 584 patients identified as possible candidates, only 199 patients met the inclusion and exclusion criteria. 177 (88.9%) of these patients were JRH patients while 22 (11.1%) were from UCLH. A summary of the excluded cases is shown in Figure 3.3 and Table 3.4.

3.6 Measurement of MRI-based tumour parameters

The measurement of MRI-based tumour parameters is central to this work. The methodology is outlined in this section, including general procedures, scan parameters, image quality assessment, volume, thickness and maximum diameter.

Table 3.4: Detailed summary of reasons for exclusion from primary study sample.

Reason	*n*	*%*	*Reason*	*n*	*%*
No MRI Recovered	*73*	*18.96%*	*Susceptibility Artefact*	*15*	*3.90%*
File Unrecoverable	*50*	*12.99%*	*Unknown Primary*	*9*	*2.34%*
Not SCC	*45*	*11.69%*	*Incomplete*	*8*	*2.08%*
Not Surgical	*39*	*10.13%*	*External Institution*	*7*	*1.82%*
Not Oral Cavity	*37*	*9.61%*	*Referred Elsewhere*	*5*	*1.30%*
Lip	*31*	*8.05%*	*Synchronous Tumour*	*5*	*1.30%*
Pre-1998	*25*	*6.49%*	*Motion Artefact*	*2*	*0.52%*
<7 days post-biopsy	*17*	*4.42%*	*Duplicate Instance*	*1*	*0.26%*
Premalignant	*16*	*4.16%*	*Total*	*385*	*100%*

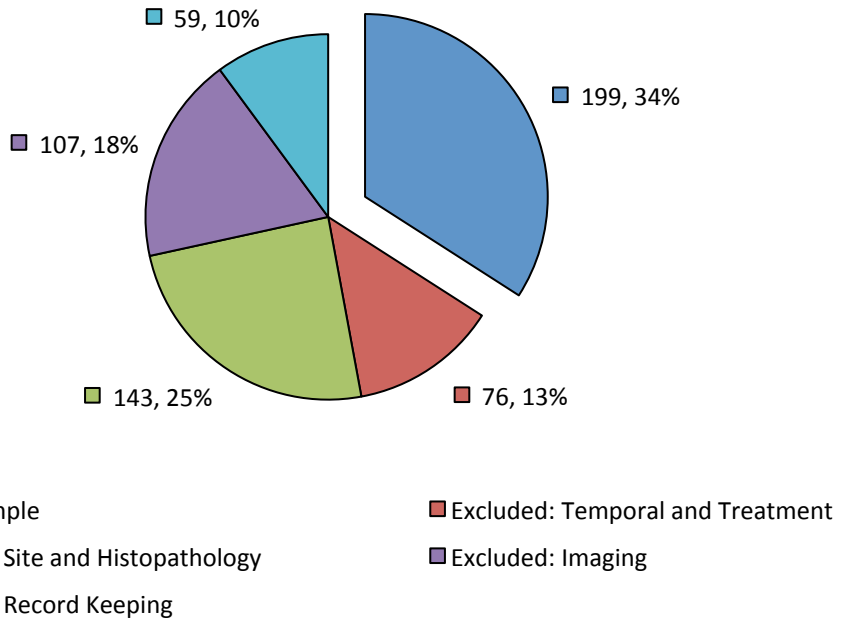


Figure 3.3: Distribution of included and excluded cases.

3.6.1 General measurement procedures

Attempts were made to avoid bias by blinding the reader to patient information. No identifying information was available to the reader apart from the unique identification number. As well the reader was not permitted to consult information contained in the patient database.

A general work flow for the measurement of tumour parameters is represented as follows:

> *A random number generator was used to determine the order in which the MRI studies were presented to the reader.*
> *Image quality and scan parameters were assessed for appropriateness.*
> *Tumour thickness was measured.*
> *Tumour volume was measured.*
> *Tumour diameter was measured.*

This protocol was repeated using a new set of random numbers to determine reading order. A time of at least one week was allowed to pass between the first and second reading protocols to minimize reader recall. The duplicate measurements were averaged to obtain the measurement value used for statistical analysis.

All measurements were carried out by a single observer, the author of this thesis. If available, the radiology report (also suitably anonymized) was consulted for guidance. In difficult cases, imaging studies were reviewed by a consultant radiologist specializing in head and neck imaging (SJG).

3.6.2 MRI scan parameters

MRI parameters varied from patient to patient. This is due to both the ten year time span of the study, as well as the use of multiple scanners at the JRH, UCLH and external referring sites. The pixel matrix size varied from 256 X 256 to 512 X 512. The field of view (FOV) varied from 20.0 cm X 20.0 cm to 30.0 cm X 30.0 cm with one case of 48.0 cm X 48.0 cm. The slice spacing varied from 0.0-2.5 mm with two extreme cases of 5.0 mm. The slice thickness varied from 3.0-6.0 mm. Echo train length varied from 0 - 20. A summary of acquisition parameters is presented in Appendix II.

A variety of pulse sequences were used to measure volume, thickness and diameter. The pulse sequences included:

T2-weighted fast-spin echo with fat saturation
T2-weighted fast-spin echo without fat saturation
T2-weighted spin echo with fat saturation
T2-weighted spin echo without fat saturation
Short tau inversion recovery (STIR)
T1-weighted fast-spin echo without fat saturation
T1-weighted spin echo with fat saturation
T1-weighted spin echo without fat saturation
Contrast enhanced T1-weighted fast-spin echo with fat saturation
Contrast enhanced T1-weighted spin echo with fat saturation

Pulse sequences were carried out in the axial, coronal and, occasionally, sagittal planes. The available pulse-sequence series of sufficient quality ranked highest on the above list was used for measurement. All available pulse sequences were reviewed to aid in locating and delineating the primary tumour. A choice was made between axial and coronal orientations based on image quality and the tumour orientation yielding the maximum number of tumour containing slices.

3.6.3 Assessment of image quality

Quality of a given series was unsatisfactory if susceptibility artefact affected the site of the primary tumour. In these cases, alternative pulse sequences less prone to susceptibility artefact were sought. If unavailable, the patient was excluded. If motion artefact obscured the tumour borders, the affected series was discarded. Alternative, less affected imaging series were sought or the patient excluded.

While susceptibility and motion artefact are the most common oral cavity imaging artefacts, a number of other artefacts are possible. Inappropriately large FOV, as well as excessive slice thickness or slice spacing were considered. Failure of fat suppression leading to poor delineation of the primary tumour was assessed. If the volume of interest did not contain the entirety of the primary tumour, no measurements were made.

3.6.4 Volume measurement

Each MRI image series was opened in ImageJ as a stack using the ''DICOM Sort'' plug-in. The image series was windowed and leveled using the ''Window/Level Adjuster'' plug-in. The ''Freehand Selections Tool'' was used to manually segment the tumour in each slice as a region of interest (ROI). The selection tool was controlled using a graphics tablet (Intuos 3 Graphics Tablet). The ''ROI Manager'' tool was used to draw and label multiple ROIs, allowing the entire tumour to be segmented prior to analysis.

If we know the number of voxels in an image series that represent tumour, and we know the dimensions of each voxel, we can calculate the volume.

Equation 3-1

$$V_v = \frac{FOV \times (Slice\ Thickness + Spacing)}{Pixel\ Matrix\ Dimensions}$$

where V_v is the voxel volume and FOV is the field of view. Since

Equation 3-2

$$V_t = N_t \times V_V$$

where V_t is the tumour volume and N_t is the total number of voxels designated as tumour, then

Equation 3-3

$$V_t = N_t \times \frac{FOV \times (Slice\ Thickness + Spacing)}{Pixel\ Matrix\ Dimensions}$$

The ''Analyze'' tool was used to calculate the area of each ROI defined during segmentation. ImageJ uses the FOV and pixel matrix dimensions contained in the DICOM metadata to calculate the area as a cm^2 value. Using Excel, the areas of the ROIs were summed and multiplied by the sum of the slice thickness and interslice spacing to yield the tumour volume. In cases where biopsy-proven SCC was not visible on coronal or axial image series, the lesion was designated as consistent with a superficial T1 lesion and given a volume of 0 cm^3.

3.6.5 Thickness measurement

The ''Straight Lines Selections Tool'' was used to draw a line, perpendicular to the mucosal surface, from the surface of the tumour to the point of deepest invasion. The ''ROI Manager'' tool was used to draw and label multiple lines in each slice. The length of each line was then calculated using the ''Analyze'' tool. ImageJ uses the FOV and pixel matrix dimensions contained in the DICOM metadata to calculate the length as a mm value. The maximum measurement value was taken as the tumour thickness. Superficial T1 lesions, invisible on MRI, were given a thickness of 0 mm.

3.6.6 Maximum diameter measurement

When measuring the tumour volume, the tumour contour was outlined in each slice. This data was recalled using the ''ROI Manager'' tool. The ''Convex Hull'' tool was used to convert the tumour outlines to convex hulls. The maximum diameter of each convex hull was measured using the ''Analyze'' tool to output the ''Feret's Diameter''. ''Feret's Diameter'' is defined by the ImageJ documentation as the longest distance between any two points along the selection boundary, also known as maximum caliper.

ImageJ uses the DICOM metadata to calculate the diameter as a mm value.

The problem with this method is that it measures the diameter only in the plane used for volume determination. Therefore, the maximum diameter in the orthogonal plane was calculated by multiplying the number of tumour containing slices by the sum of the slice thickness and slice spacing.

Equation 3-4

$$D_o = n_{slices} \times (Slice\ thickness + Spacing)$$

where D_o is the maximum diameter in the orthogonal plane and n_{slices} is the number of tumour containing slices. The largest measured value of either method was taken to be the maximum diameter. Superficial T1 lesions, invisible on MRI, were given a diameter of 0 mm.

3.7 Prognostic and predictive factors

The background for the prognostic and predictive factors were discussed in Chapter 2. This section outlines the procedure for gathering data for each individual patient and tumour-related factor.

3.7.1 Patient factors

3.7.1.1 Age and gender

Age was defined as the interval between the date of birth and primary surgical resection. The gender was recorded from the patient record.

3.7.1.2 Alcohol

Alcohol consumption was typically gathered from the social history taken at either the initial consultation or pre-surgical admission. In some cases, the information was available in correspondence with other healthcare professionals.

Significant alcohol consumption was defined as drinking in excess of 21 units of alcohol per week or if there was a subjective indication of excessive alcohol consumption. Though 21 units per week is the recommended upper limit of consumption for males (the upper limit for females being 14 units) the higher limit was universally adopted to simplify data acquisition [189]. This is similar to the methods of Mayne et al. who used 21 units [122].

3.7.1.3 Smoking

Information about smoking was obtained from the same source as alcohol consumption. A smoker was defined as a current smoker or a smoker in the ten year period prior to surgical resection. No quantitative metric of cigarette smoking, such as pack-years, was used to define a positive smoking history because such detailed information was typically un

3.7.1.4 Adult comorbidity evaluation - 27

The ACE-27 comorbidity scale, detailed in Appendix I, was used to define the level of patient comorbidity. All information was found in the patient record. The sources of information used within

the patient records were:

The patient history taken at initial evaluation and/or biopsy
The patient history taken during pre-surgical admission and workup
Vital information taken by nursing staff at the time of pre-surgical admission
Referral documentation
Patient-surgeon correspondence
Documentation of specialist consultation

3.7.1.5 Institution

In this study the two participating institutions were the John Radcliffe Hospital (JRH) and University College London Hospital (UCLH).

3.7.2 Tumour factors: clinical

3.7.2.1 Oral cavity subsite

The anatomical subsites of the oral cavity are defined by the AJCC [190]. In cases where the tumour occupied multiple subsites, a site of origin was chosen based on information provided in the patient record and the area of greatest involvement.

The floor, lower alveolus and retromolar trigone were combined to form the group "lower oral cavity", while the hard palate, upper alveolus and buccal mucosa were combined to form the group "upper oral cavity". Tumours of the oral tongue remained in their own group. The groupings are summarized in Table 3.5.

Grouping was necessary to avoid the small case numbers that occur when the study sample is divided by the large number of subsites. This is a difficult task. SCC arising at each oral cavity subsite has a distinct natural history, though recent work indicates that there is a common molecular thread [191]. Regional metastasis and prognosis also varies with subsite.

The subsites were divided into upper and lower oral cavity because, anatomically, this is a sensible division that provides a reasonably similar sample size in each category. The implications of this

grouping scheme are discussed Chapter 5.

3.7.2.2 Location

Tumours were classified as originating on the right, left or midline of the oral cavity. If tumours involved the midline but were thought to be lateral in origin, they were classified according to origin.

3.7.2.3 Growth Pattern

The growth pattern was determined from subjective descriptions in the patient record. Tumours were designated as exophytic if it was described as such, if it was described in such a way that there was a high probability the tumour was exophytic or if there was a diagram or photograph that showed the tumour as exophytic.

No distinction was made between endophytic or ulcerative tumours. Both growth patterns were referred to as endophytic. Tumours were endophytic if they were described as endophytic or ulcerative in the patient record, if they were described in such a way that there was a high probability the tumour was endophytic or ulcerative, or if there was a diagram or photograph that showed the tumour as endophytic or ulcerative.

Table 3.5: Grouping of oral cavity subsites.

Grouping	*Oral cavity subsite*
Upper oral cavity	*Hard palate*
	Upper alveolus
	Buccal mucosa
Lower oral cavity	*Retromolar trigone*
	Lower alveolus
	Floor
Oral tongue	*Oral tongue*

3.7.3 Tumour factors: staging

3.7.3.1 TNM

The TNM-classification was based on the AJCC's Cancer Staging Manual (6th Edition) [13].

3.7.3.2 T-stage

The clinical T (cT) was initially determined from notes taken during the initial consultation and examination. The cT was then refined by staging MRI and a composite cT recorded. In some cases the cT was determined from correspondence. Further, in some cases of referral from other institutions, the results of imaging were available prior to clinical examination. Criteria for cT4 tumours were also evaluated by clinical examination. However, MRI and CT played the more prominent role in the evaluation of bone and critical soft tissue invasion. As a result, only radiology-determined cT and a composite cT could be evaluated. A cT based on only clinical examination could not be determined.

The pathological T (pT) was found in the pathology report following primary surgical resection. A general procedure for pathological examination is discussed in 3.7.4.

3.7.3.3 N-stage

The clinical N (cN) was based on documentation of staging clinical examination or radiologist reports. Information sources were the same as those for cT, with similar associated problems.

CLNM was evaluated on MRI to augment manual palpation. In the majority of cases, CLNM status was obtained from the radiologist report. However, in some cases where the radiologist report was unavailable, the CLNM status was evaluated by direct review of the staging MRI study by the author with guidance from SJG, a consultant radiologist, when required.

The CLNM status was reported as negative, positive or equivocal. In cases where there were enlarged lymph nodes but in which there were mitigating factors, such as signs of global cervical lymph node inflammation, the finding was reported as equivocal. CLNM status was reported as positive when one or more criteria for metastasis led to a high suspicion of the presence of disease.

The pN was determined from the pathology report following primary surgical resection. In most cases, the pathology report listed a standardized dataset, including the number of lymph nodes dissected, number and the level of metastatic nodes, diameter of the largest metastatic deposit, extracapsular spread (ECS) and the number and level of any nodes with ECS. A general procedure for pathological analysis is discussed in Section 3.7.4.

3.7.3.4 M-stage

No patient was found to have distant involvement at the time of primary surgery. Distant involvement, when evaluated, was investigated using chest X-ray or chest CT. In cases where distant metastasis was not evaluated, the case was designated cMX, however for the purposes of the final clinical stage reported, cMX patients were treated as cM0.

3.7.3.5 Extracapsular spread

The clinical ECS was determined from MRI. Typically, ECS was found in the radiologist report. When the radiology report was unavailable, ECS was determined by the author with guidance from SJG when required. Pathological ECS was determined from pathology reports. A general procedure for pathological analysis is discussed in Section 3.7.4.

3.7.3.6 Bone invasion

It is the usual local practice to examine bone invasion on both MRI and CT, except in cases of localized SCC of the tongue. Co-registration of MRI and CT is carried out at the JRH to evaluate cases with limited invasion. Information regarding bone invasion was available from radiologist reports. In cases in which the radiology report was unavailable, the CT and MRI studies were inspected by the author with guidance from SJG when required.

3.7.3.7 Midline involvement

Midline invasion was evaluated on MRI. Tumours not approaching the midline and those crossing the midline were classified as such. However, following the convention found in radiologist reports a third category, abutting the midline, was introduced. These tumours approached very near to, or

contacted, the midline but did not invade contralateral tissue. This was especially apparent in the oral tongue, where the midline septum served as a visible guide.

In the majority of cases, lateral tumours did not approach the midline. Many large tumours had obvious contralateral components. In difficult cases, the radiologist's interpretation was often found in the report. Failing this, the MRI study was inspected by this author with guidance from SJG.

3.7.4 Tumour factors: histopathology

The preparation and evaluation of pathology specimens was based on recommendations by the Royal College of Pathologists' Datasets for Histopathology Reports on Head and Neck Carcinomas and Salivary Neoplasms [192]. This is the regular practice at the institutions involved in this study (Personal correspondence, Dr. Ketan Shah).

3.7.4.1 Preparation of the primary tumour

Specimens are oriented by the surgeon and pinned or sutured to cork or polystyrene boards. Margins were identified by sutures or metal tags and fixed for 24-48 hours in a formalin-based solution. Specimens were cut into 5 mm slices that highlight the mucosal and deep resection margins. If bone was included in the specimen it was decalcified before examination.

The blocks of tissue to be examined were selected as follows:

> *At least one block per 10 mm of tumour diameter, including one selected to show the maximum depth of penetration. If the tumour is less than 10 mm in diameter, the whole tumour was examined.*
>
> *Blocks of defined tissue margins*
>
> *Non-neoplastic mucosa*
>
> *Bone surgical margins*
>
> *Bone, if involved by tumour*

Several core items are included in each pathology report and are outlined in Table 3.6.

Table 3.6: Core items to be reported for examination of the primary tumour. The Royal College of Pathologists recommend that the following tumour characteristics are reported for each tumour of the oral cavity examined [192].

Maximum diameter and depth of invasion
The maximum diameter was measured macroscopically, unless it was not visible, in which case microscopic measurements were used. The depth of invasion was measured microscopically from a reconstructed mucosal surface to the point of maximum invasion.
Histological type
Papillary, verrucous, basaloid, adenosquamous and spindle cell variants, if present, were recognized and reported.
Tumour differentiation
Grade was based on the degree of resemblance to normal mucosa and follow the conventions of the WHO classification. Using medium magnification, the most aggressive area was graded as either well, moderately or poorly differentiated.
Infiltrative front
The pattern of infiltration at the deep margin was classified as either cohesive (pushing) or non-cohesive.
Mucosal and deep surgical margins
The distance from the tumour to the margin was measured histologically. The closest measurement was used to determine if the margin is clear (>5 mm), close (1-5 mm) or involved (<1 mm).
Lymphovascular invasion
The presence of tumour cells within an endothelial-lined space. Endothelial-lined space can refer to venous or lymphatic vessels.
Perineural invasion
The presence of invasion of the perineural space ahead of the invasive front, not intratumoural perineural invasion, was recorded.
Bone invasion
Bone involvement was described as non-invasive cortical erosion or as diffuse infiltration of medullary intratrebecular and perineural tissues.
Dysplasia
The presence of severe dysplasia or carcinoma in situ near the tumour or at the resection margins was recorded.
pT
The pT was recorded based on the guidelines of the AJCC [13].

3.7.4.2 Preparation of neck dissection specimens

Neck dissections specimens were oriented by the surgeon and pinned or sutured to a cork or polystyrene board. Critical margins were indicated and the level of nodal groups marked with sutures or metal clips. Specimens were fixed a formaldehyde-based solution for 24-48 hours.

Each individual lymph node was identified and dissected with attached pericapsular adipose tissue. Large nodes were bisected or sliced. If there was obvious tumour, the slice with the more extensive disease was processed to document ECS. Negative nodes were processed in their entirety. Small or flat nodes were processed whole. Core data items were outlined in Table 3.7.

Table 3.7: Core items to be reported for examination of the neck dissection. The Royal College of Pathologists recommend that the following neck dissection characteristics are reported for each specimen examined. [192].

Nodes
The number identified and the number of nodes with metastatic disease were recorded for each level present in the resected specimen. The diameter of the largest metastatic deposit was recorded. This was not equivalent to the size of the largest node. This was an important factor for pN staging.
Extracapsular spread
The presence of tumour spread beyond the lymph node capsule into the perinodal fat was assessed histologically and the levels involved identified. Extensive ECS was evaluated for extension into adjacent structures. If the presence of ECS was uncertain, pathologists errored on the side of caution and recorded that ECS was present.
Micrometastasis
The presence as well as the number and level of involved nodes with metastatic deposits less than 2 mm in diameter were recorded. The presence of isolated clumps of tumour cells visible only on immunohistochemical staining was reported.
Matting of lymph nodes
When the fusion or matting of lymph nodes was present, the levels involved, the maximum diameter of the fused deposit and an estimation of the number of nodes involved were recorded.
Isolated tumour nodules
Isolated tumour nodules found in the lymphatic drainage distribution were regarded as nodal metastasis with the exception of those nodules that were within 10 mm of the primary tumour. In the absence of residual lymphatic tissue, these nodules were recorded as discontinuous extensions of the primary tumour.
pN
The pN was recorded based on the guidelines of the AJCC [13].

3.8 Conclusions

This chapter is meant to provide the reader with detailed information about the methods and procedures that are important to all aspects of this work. In future chapters, methodology specific to each component of the work, not covered in the general methodology, will be given. The reader is

invited to return to this chapter and use it as a reference when necessary. Where helpful, the text will refer the reader to relevant sections of the general methodology to aid in understanding of subsequent chapters.

The focus of the thesis now shifts into the main experimental thrust of the work. The thesis now follows three parallel paths that eventually combine into a general discussion and conclusions. The following two chapters take an in-depth look at the role of depth of tumour invasion, the controversy surrounding the definition of an optimum cut-point for TNM staging, and the role of statistical methodology in cut-point variability.

4 Statistical methods: a critical review

4.1 Introduction

Depth of invasion has been championed as a replacement of maximum diameter in the TMN staging criteria for the oral cavity. Consequently, a great deal of effort has been put into trying to find the "optimal cut-point" with which to classify patients into low and high risk groups.

Current efforts to establish the depth of invasion as a predictor of OCLNM and survival have led to the introduction of a wide variety of statistically significant cut-point values in the literature. The result has been widespread disagreement as to the optimal value, with a number of candidates emerging. However, when one considers the process of categorizing continuous data it is apparent that the cost of simplification is a loss of information. Furthermore, categorizing continuous data without proper consideration for appropriate statistical rigor can lead to results with underestimation of p-values and overestimation of effect sizes.

Taking this into consideration, it is worthwhile to consider the use of cut-point values in the OCSCC literature, with emphasis on statistical methods. The aim of this review of statistical methods is not to criticize or discredit the work of other researchers in this field. It is to draw attention to this particular statistical issue that may be present in the literature with the intent of guiding future work, including the work in this thesis.

4.1.1 Chapter goals

The goals of this chapter are:

1. *To establish the emergent role of depth of invasion as a prognostic factor in oral cancer and highlight the controversy associated with establishing an optimal cut-point.*

2. *To outline the practice of dichotomizing continuous variables, including possible motivations and the inherent problems caused by information loss and multiple testing.*

3. *To review the depth of invasion literature with the intent of establishing the extent of cut-point variability, the most commonly used statistical methods, and poor statistical practice.*

4. *To use statistical simulation to quantify the problems with cut-point variability and*

inflated p-values associated with multiple testing and "optimum p-value" selection.

5. *To establish statistical methodology as a source of cut-point variability and to suggest solutions for future study.*

4.1.2 Organization of sections

The chapter begins with background about the elusive depth of invasion cut-point in OCSCC. A brief outline of depth of invasion as a prognostic and predictive factor is given. Sources of cut-point variability are outlined. The background goes on to expand upon the role of statistical methodology in cut-point variability, focusing on the problems associated with the dichotomization of continuous variables.

At this point the chapter diverges into two parallel sections. In the first section, the methodology and results of an extensive review of the depth of invasion literature are presented. The review focuses on proposed depth of invasion cut-points and the appropriateness of associated statistical methods. A case is made to support the hypothesis that questionable statistical practices are a major source of cut-point variability. The discussion is reserved for a later section.

In parallel to the cut-point literature review is a statistical simulation study. The structure of each simulation is outlined in the first section. The simulation study focuses on the danger of high numbers of falsely-positive tests associated with continuous variable dichotomization in a manner specific to oral cavity cancer depth of invasion. The results of each simulation are outlined and the discussion reserved for the following section.

The literature review and simulation study are brought together in a unified discussion. Implications for the interpretation of the current oral cancer literature and future study design are outlined. The chapter closes with a brief summary and conclusions.

4.2 Background

This first part of the background will establish depth of invasion as an emergent predictive and

prognostic factor in oral cavity SCC. The remainder of the background discusses the rationale for continuous variable categorization and the problems that can arise from the use of the "lowest p-value" method.

4.2.1 The elusive cut-point in oral cavity cancer management

A 2005 review by Pentenero et al. found that there is widespread disagreement about the existence of a single cut-point to guide OCSCC prognosis and management [193]. The authors noted that values reported by publications ranged from 1.5-10 mm and concluded that further study was needed [91, 194-195].

Oral cavity heterogeneity was suggested as one reason for variability. Publications used a range of oral cavity sites or groups of oral cavity sites. Theoretically, risk of metastasis and survival could be related to oral cavity subsite as these cancers have different natural histories [191]. Further, at some sites, such as the retromolar trigone, critical structures like the mandible are relatively close to the surface. There may also be some site-related variation in lymphatic vessel density.

Publications also reported using a number of different cT and cN stages as inclusion criteria. A sample composed of T1-4 tumours would result in larger cut-point than samples limited to early T1 and T2 tumours.

Measurement techniques are another source of cut-point variability. The terms ''depth of invasion'' and ''thickness'' were often used as synonyms, though, as Moore et al. contend [196], they are separate entities. Depth relies on the use of a reconstructed mucosal border. This reconstruction is somewhat subjective and techniques adopted by pathologists vary. In some cases, a reconstructed basement membrane is used instead of a mucosal border. Tumour thickness measurement is more objective; however subjective elements persist, for example in the selection of the point of deepest invasion.

All of these factors play some role in the range of cut-points that are reported in the oral cancer

literature. However, other factors must be considered.

In a review of the breast cancer literature concerning the prognostic factor ''DNA S-phase fraction'' (SPF), Altman et al. explored the statistical implications of cut-point selection [197]. The group concentrated on the use of the ''optimal cut-point'' technique, a procedure by which a series of potential cut-points are tested using χ^2 or log rank methods. The cut-point with the lowest p-value is taken to be ''optimal''. While this approach seems intuitive, problems arise from multiple testing. Its use could lead to a 10-fold increase in false-positive rate. In addition, the cut-point is highly data dependent and would vary widely from sample to sample.

While the extent of the ''optimal cut-point'' technique in the OCSCC literature is unknown, it is worthwhile to investigate this potential source of variability. If this is a contributing factor, no amount of further testing with standard samples and measurement techniques can resolve these issues without statistical techniques first being addressed.

4.2.2 Why do we dichotomize?

There are several reasons why the dichotomization of continuous data is commonplace in the medical literature.

4.2.2.1 Clinical simplicity

The categorization of patients into distinct groups provides a straightforward method of directing clinical decisions. Patients are given labels such as "obese", "hypertensive" or "high cholesterol" based on continuous data.

4.2.2.2 Statistical simplicity

When continuous data is dichotomized statistical analysis is simplified. Troublesome concerns about the non-linearity of the continuous dataset can be ignored and the test of choice is rarely complex.

By comparison, analysis of continuous data can be complex. Non-linear data often must be transformed to meet normality assumptions required by many statistical tests. Outliers must be

considered carefully as their potential to skew results is enhanced. Analysis requires mathematically complex regression methods that yield results that are less easily interpreted and more difficult to report.

4.2.2.3 Congruity with previous work

The use of continuous data dichotomization is extremely common [198]. In a particular field, it may be that the majority of published work uses this approach. It is only natural for researchers to adopt techniques similar to their colleagues to allow for scientific exchange and comparison. In this way methods of statistical analysis become self-perpetuating.

4.2.2.4 Compensation for imprecise measurement

MacCallum et al. reported that some researchers feel that analysis of imprecisely measured data is more reliable when the information is dichotomized into high and low measurement groups [199].

4.2.3 Problems

Dichotomization of continuous variables results in a loss of information. Consider a situation in which the risk of a disease increases linearly with age. If a cut-point is used to separate patients into a high-risk and low-risk group, any information about the variation in risk within each risk group is lost. We are left with a situation in which a patient slightly above the cut-point is deemed to be at markedly higher risk than a patient slightly below the cut-point.

The consequence is a decrease in statistical power. Indeed, Lagakos et al. showed that dichotomizing a normally distributed continuous variable using the median as the cut-point results in a 35% loss of statistical power [200]. When multiple cut-points are used to divide the data, the categories more closely approximate the linear increase in risk with age. However information is still lost. This is a particular problem in studies with small sample sizes that have limited statistical power to begin with, as is the case with the oral cancer prognostic factor literature.

There are several ways in which investigators determine the cut-point used. *A priori* cut-points may

be based on previously reported literature values. Alternatively, the median of the continuous data may be used. These approaches can lead to the reporting of widely varying cut-point variables and results are not easily compared [198].

Investigators have reported using an ''optimal cut-point'' technique for dichotomizing their data. This approach involves testing every possible cut-point then choosing the value that minimizes the p-value. Practically, this means that authors usually choose several potential cut-points from the range of the continuous variable. This sets up a situation in which the problem of multiple testing can produce falsely-significant results.

Without some sort of p-value correction to account for multiple testing, the ''optimal cut-point'' technique can lead to over-optimistic results. In their review of the breast cancer literature, Altman et al. reported that the "optimum cut-point" approach could lead to a 10-fold increase in false-positives [197]. In addition, the cut-point selected by this method is highly data dependent and varies widely from sample to sample. The authors also report that the ''optimal cut-point'' approach leads to considerable overestimation of the effect size of prognostic factors.

Hollander et al. illustrated the problems of the ''optimal cut-point'' approach using a dataset of 686 node-positive breast cancer patients [201]. Patient age and 5-years event-free survival (EFS) were recorded. Independence between the two variables was achieved by randomly reassigning the age values. The ability of age to predict EFS was then considered using the log rank test by dichotomizing data using the "optimal cut-point" approach. This procedure was repeated 100 times. 32% of results had a p-value less than 0.05 and were classified as significant. Almost every possible age value was selected as the "optimal cut-point" in one or more iterations of their simulation.

4.2.4 Solutions

Numerous ways have been proposed to account for the bias introduced by multiple testing. In the Bonferroni correction, the p-value is divided by the number of statistical tests to yield an adjusted value which is then used to assess significance. This method of adjustment is noted to be

conservative, especially when large numbers of cut-points have been evaluated [202].

An alternate approach was outlined in a 1994 paper by Altman et al. [197], based on previous theoretical work [203-204]. A formula was proposed to calculate a corrected value of the chi-square statistic for situations involving large sample sizes in which multiple cut-points are being evaluated. The corrected p-value, P_{cor}, is obtained as follows:

Equation 4-1

$$P_{cor} = \varphi z(z - \frac{1}{z}) \ln \frac{(1-\varepsilon)^2}{\varepsilon^2} + \frac{4\varphi z}{z}$$

where φ denotes the probability density function and z is the $(1 - P_{min}/2)$ quantile of the standard normal distribution. ε refers to the proportion of the largest and smallest continuous data points that are not evaluated as cut points.

The above equation is simple to calculate, but requires the use of the probability density function of the standard normal distribution. For convenience, the formula can be approximated by:

Equation 4-2

$$P_{cor} = -1.63P_{\min}(1 + 2.35 \ln P_{min}) \text{ for } \varepsilon = 10\%$$

Equation 4-3

$$P_{cor} = -3.13P_{\min}(1 + 1.65 \ln P_{min}) \text{ for } \varepsilon = 5\%$$

These equations can be applied to existing studies to evaluate the significance of the results without the influence of multiple testing.

4.3 Methods: literature review

The next step in establishing statistical methodology as a source of cut-point variability is to review the depth of invasion literature with the intent of assessing statistical practice. In this section, key characteristics of papers that study the role of depth of invasion in cervical lymph node metastasis prediction and survival are collated.

4.3.1 Literature review: thickness and depth of invasion

A review of the literature concerning the relationship of depth and thickness CLNM and survival in OCSCC was carried out. Search criteria were defined using the methods of a review by Pentenero et al. [193]. A search of the PubMed database (National Institutes of Health, Bethesda, USA) was conducted using the following parameters:

> *((((''tumor thickness'' [All Fields] or ''tumour thickness'' [All Fields] not ''full-thickness'' [All Fields]) or depth [All Fields]) and (''mouth neoplasms'' [MeSH Terms] or ''oral cancer'' [All Fields])) and (''human'' [MeSH Terms] or ''hominidae'' [MeSH Terms] or ''Human'' [MeSH Terms]) and (''1984'' [PDAT]: ''3000'' [PDAT]) and (''human'' [MeSH Terms] or ''hominidae'' [MeSH Terms] or ''Human'' [MeSH Terms]) and (''1984'' [PDAT]: ''3000'' [PDAT]).*

Papers identified by the search criteria were screened using the title and abstract for reference to depth and/or thickness, as well as CLNM and/or survival. Only papers evaluating SCC of the oral cavity or an oral cavity subsite were included. Papers not performing any type of statistical analysis were excluded, as were reviews. Additional studies identified through the references of evaluated studies were included. Papers meeting the inclusion and exclusion criteria were reviewed and the following parameters recorded:

Anatomic locations included
Primary treatment
Sample size
cT and cN categories included
Measurement: depth or thickness as reported by the author
Outcomes studied: ACS, DFS, DRS, CLNM, OCLNM
Statistical tests undertaken
Cut-point value used (if applicable)
Method used to derive cut-point value (if applicable)
p-value

4.3.2 Defining sample, statistical test and cut-point values

A distinct sample was defined as a group of patients unique with respect to one or more of sample origin, sample size, anatomical site, treatment, cT or cN. A single publication could have multiple samples.

A statistical test was defined as the evaluation of the relationship between a single predictor variable and a single outcome that is distinct in terms of sample or statistical methodology. For example, a paper evaluating the relationship between tumour depth and CLNM including the entire oral cavity as well as a subsample including only the oral tongue would constitute two statistical tests. For papers evaluating both depth and thickness, these were recorded as distinct statistical tests.

The median and range of the sample sizes for each outcome category was calculated using values recorded for each distinct sample. For the median and range of the cut-point values, a value was recorded for each distinct statistical test. If a paper tested multiple thickness or depth cut-points using the same sample, then the preferred cut-point was recorded based on the discussion and conclusion sections. If no preference was expressed, then a representative cut-point was chosen based on the lowest p-value. In cases where three or more categories were generated using multiple cut-points, a value was recorded for each.

4.3.3 p-value correction

The Altman et al. p-value correction equation was applied to statistical tests assuming ε values of 10%. The correction equation was applied to all χ^2, Fisher's exact and log rank tests using cut-point values for the depth or thickness in which the method of cut-point selection was arbitrarily selected. For the purposes of this study, when logistic regression or Cox proportional hazard testing were done using dichotomized depth or thickness, these were considered equivalent to χ^2 and log rank testing.

A cut-point value was considered arbitrarily selected if:

The cut-point value tested was not defined in the materials and methods.
The cut-point value was not defined in the results or discussion of the paper based on:
A mean, median or percentile value.
Previously published depth of invasion or thickness studies.
Sensitivity and specificity or ROC analysis.
The cut-point value was defined but no reason for the choice was given.
The author described using multiple cut-points to dichotomize thickness data.

p-value correction was not applied if:

Thickness or depth was treated as continuous data and not categorized.
The cut-point value tested was expressly defined in the materials and methods and derived from:
A mean, median or percentile value.
Previously published depth of invasion or thickness studies.
Multiple categories were defined using multiple cut-points.
A statistical test other than χ^2, Fisher's exact, log rank, logistic regression or Cox proportional hazards was applied to dichotomized data.

4.4 Results: literature review

In the previous section, methodology was established for the review of the depth literature. Key concepts were defined and a procedure for p-value correction established. In this section, sample size, proposed cut-points and statistical methodology of each paper are summarized and presented. Survival and CLNM are discussed separately. For both survival and CLNM, the results are presented in tabular form where possible. Otherwise, figures or text summaries are used. Tables and figures are found at the end of the section.

4.4.1 Literature review: depth of invasion

The review returned a total of 217 papers. 66 papers meeting the inclusion and exclusion criteria were identified with 137 statistical tests. For the study of survival, 25 papers conducting 52 statistical tests on 30 distinct samples were identified. 51 papers conducting 85 statistical tests on 59 samples studied CLNM. The data contained in each publication grouped according to the outcome studied is summarized in Appendix III.

4.4.2 Survival: ACS, DRS and DFS

Survival studies were very heterogeneous in both sample size and composition. The distribution is summarized in Figure 4.1. Similarly, a variety of TNM classifications were used as inclusion criteria, summarized in Table 4.1. Surgical resection was the primary treatment in all studies except for 4 that combined surgery and non-surgical therapy. Sample size varied from 20 to 827 with a median of 86. 9 (30%) of the samples had less than 50 while 8 samples (27%) were greater than 100. In the majority of cases depth or thickness was studied on histopathology except in 1 study (2.5%) using ultrasound.

A number of statistical methods were used to analyze measurement data, shown in Figure 4.1. The median and range of the cut-points for each sample are given in

Table 4.2. Table 4.3 classifies the statistical tests, before and after p-value correction, as significant or non-significant based on a threshold value of p = 0.05.

In the majority of studies, authors were unclear about their method of choosing a particular cut-point. A minority of publications evaluated cut-point values derived from previously published studies, the sensitivity and specificity or the minimum p-value. One study used quartiles as cut-points. No studies used the median or mean.

4.4.3 Cervical lymph node metastasis

The CLNM papers were also heterogeneous in sample size and composition, summarized in Figure 4.3. Similarly, a variety of TNM combinations were used as inclusion criteria, summarized in Table 4.1. Surgical resection was the primary treatment in all studies except for 1 (2%) in which patients received brachytherapy. The sample size varied from 20 to 184 people with a median of 51. 29 samples (49.1%) were composed of less than 50 people while 10 samples (16.9%) were greater than 100. In the majority of cases depth or thickness was studied on histopathology except in 1 study (2%) using ultrasound and 2 (4%) studies using MRI.

A number of statistical methods were used to analyze measurement data in the publications. The tests used are listed in Figure 4.4. The median and range cut-points are given in

Table 4.2. Table 4.3 classifies the continuous and categorized statistical tests before and after p-value correction as significant or non-significant based on a threshold value of p = 0.05.

In the majority of publications, authors gave no justification for the cut-point used. A minority of publications evaluated cut-point values derived from previously published studies, sensitivity and specificity or the minimum p-value approach.

Table 4.1: Combinations of T-staging and N-staging criteria by outcome. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival; CLNM: cervical lymph node metastasis; OCLNM: occult cervical lymph node metastasis.

T	*N*	*ACS*	*DRS*	*DFS*	*CLNM*	*OCLNM*
T1	*N0*	*1*	*-*	*-*	*-*	*1*
T1	*N0-3*	*-*	*-*	*-*	*1*	*-*
T1-2	*N0*	*4*	*1*	*6*	*-*	*17*
T1-2	*N0-2*	*-*	*-*	*-*	*2*	*-*
T1-2	*N0-3*	*-*	*-*	*2*	*2*	*-*
T1-3	*N0*	*1*	*-*	*-*	*-*	*2*
T1-3	*N0-2*	*-*	*-*	*-*	*2*	*-*
T1-3	*N0-3*	*-*	*-*	*1*	*1*	*-*
T1-3	*N1*	*-*	*-*	*1*	*-*	*-*
T1-3	*N1-2*	*-*	*-*	*1*	*-*	*-*
T1-4	*N0*	*-*	*-*	*-*	*-*	*4*
T1-4	*N0-1*	*-*	*-*	*-*	*1*	*-*
T1-4	*N0-2*	*-*	*-*	*-*	*1*	*-*
T1-4	*N0-3*	*3*	*1*	*5*	*23*	*-*
T2	*N0-3*	*1*	*-*	*-*	*1*	*-*
T2-4	*N0-3*	*-*	*-*	*-*	*1*	*-*
T3	*N0*	*1*	*-*	*1*	*-*	*-*
T4	*N0*	*1*	*-*	*1*	*-*	*-*

Table 4.2: The median and range of cut-point values for each outcome of interest. ACS: all-cause survival; CLNM: cervical lymph node metastasis; DFS: disease-free survival; DRS: disease-related survival; OCLNM: occult cervical lymph node metastasis.

Cut-point (mm)					
	ACS	*DRS*	*DFS*	*CLNM*	*OCLNM*
	5.0 (3.0 - 10.0)	*5.0 (2.0 - 10.0)*	*5.0 (3.0 - 10.0)*	*4.75 (1.5 - 9.0)*	*4.0 (1.0 - 7.0)*
Sample Size					
	ACS	*DRS*	*DFS*	*CLNM*	*OCLNM*
	87.5 (44 - 827)	*85 (20 - 232)*	*89 (20 - 827)*	*50 (20 - 184)*	*50.5 (25 - 172)*

Table 4.3: The number of significant (S) and non-significant (NS) statistical tests for each outcome. The results for continuous statistical tests, cut-point tests without correction and cut-point tests after correction are shown. ACS: all-cause survival; CLNM: cervical lymph node metastasis; DFS: disease-free survival; DRS: disease-related survival; NS: not significant; OCLNM: occult cervical lymph node metastasis; S: significant

	ACS		*DRS*		*DFS*		*CLNM*		*OCLNM*	
	S	*NS*	*S*	*NS*	*S*	*NS*	*S*	*NS*	*S*	*NS*
Continuous	*1*	*0*	*1*	*1*	*1*	*1*	*0*	*0*	*0*	*0*
Cut-point	*14*	*5*	*8*	*2*	*13*	*4*	*0*	*0*	*0*	*0*
Corrected	*9*	*10*	*5*	*5*	*7*	*10*	*0*	*0*	*0*	*0*

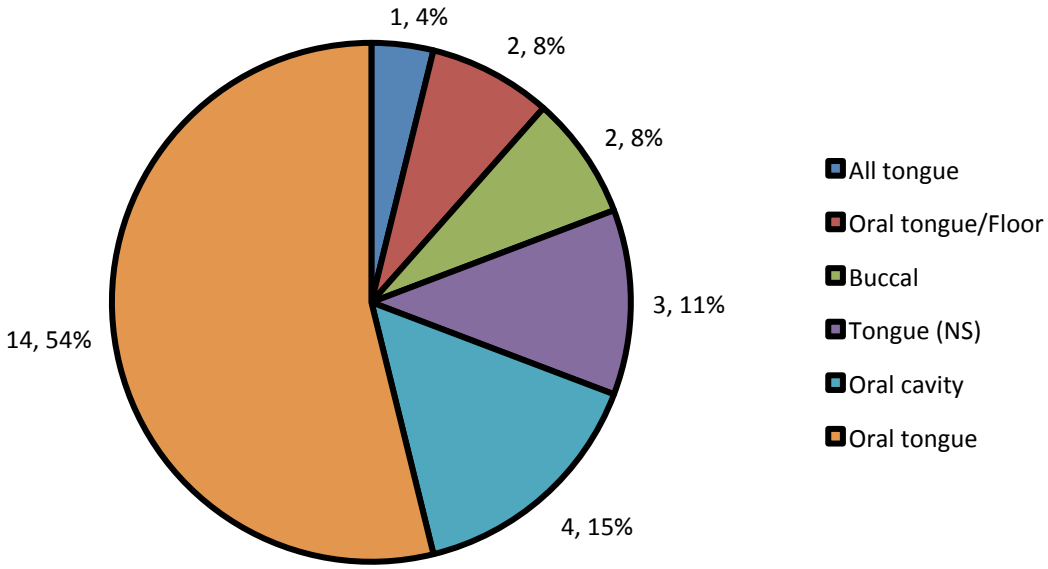


Figure 4.1: Pie chart of the distribution of anatomic sites for survival studies. Each unit represents a single publication studying the given site. NS: tongue not otherwise specified.

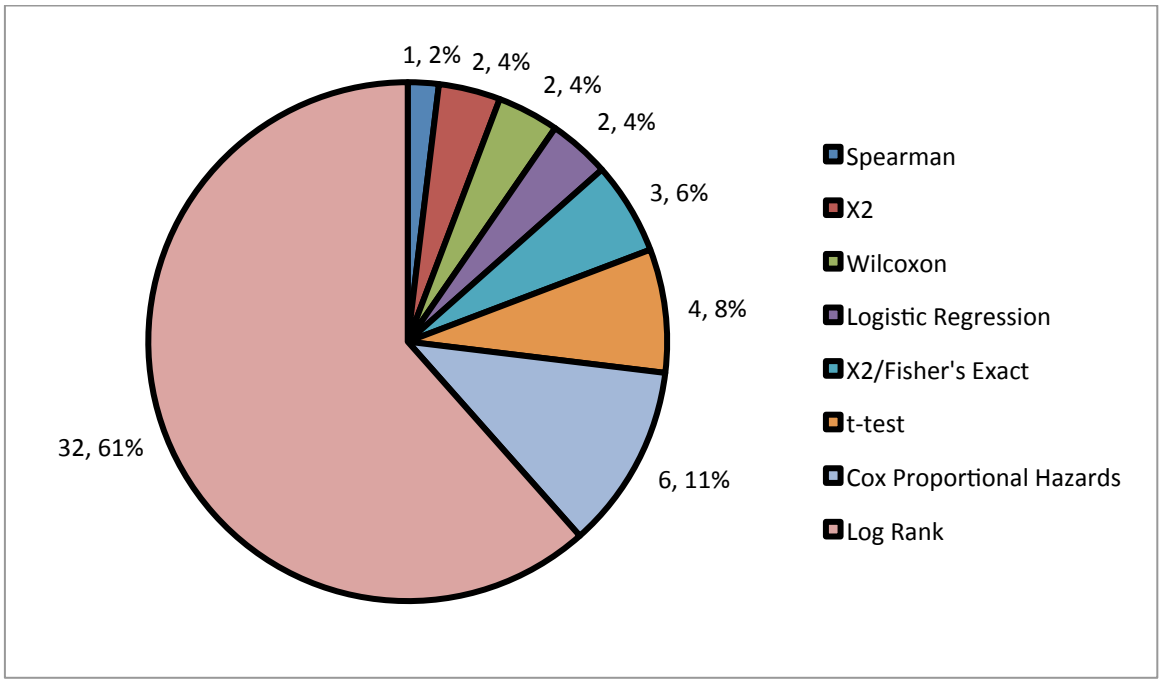


Figure 4.2: Pie chart of the distribution of statistical tests for survival studies. Each unit represents a single publication studying the given site.

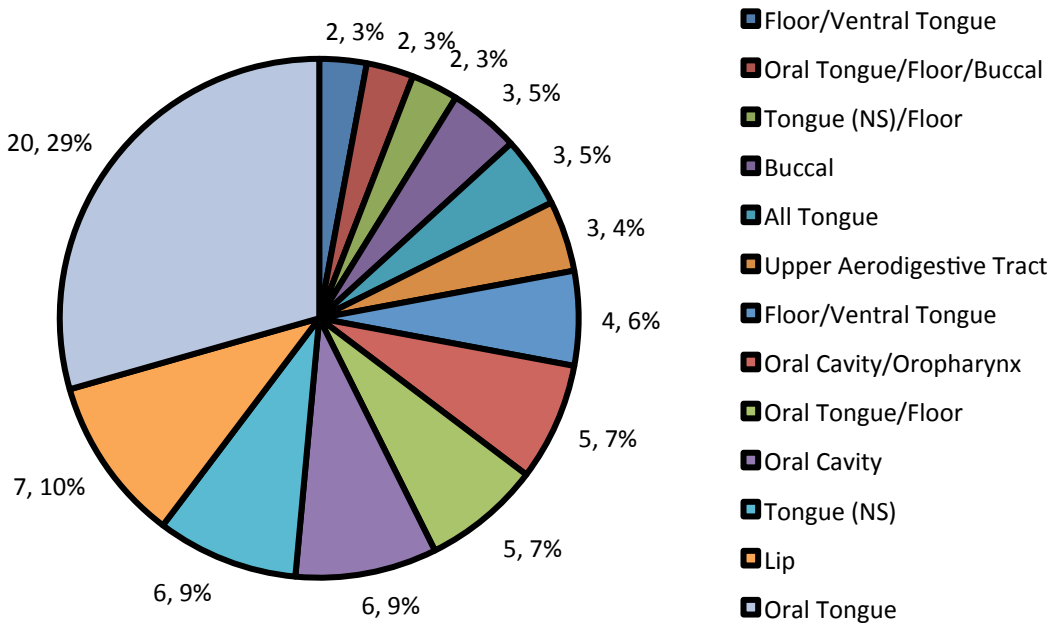


Figure 4.3: Pie chart of the distribution of anatomic sites for CLNM studies. Each unit represents a single publication studying the given site. NS: tongue not otherwise specified.

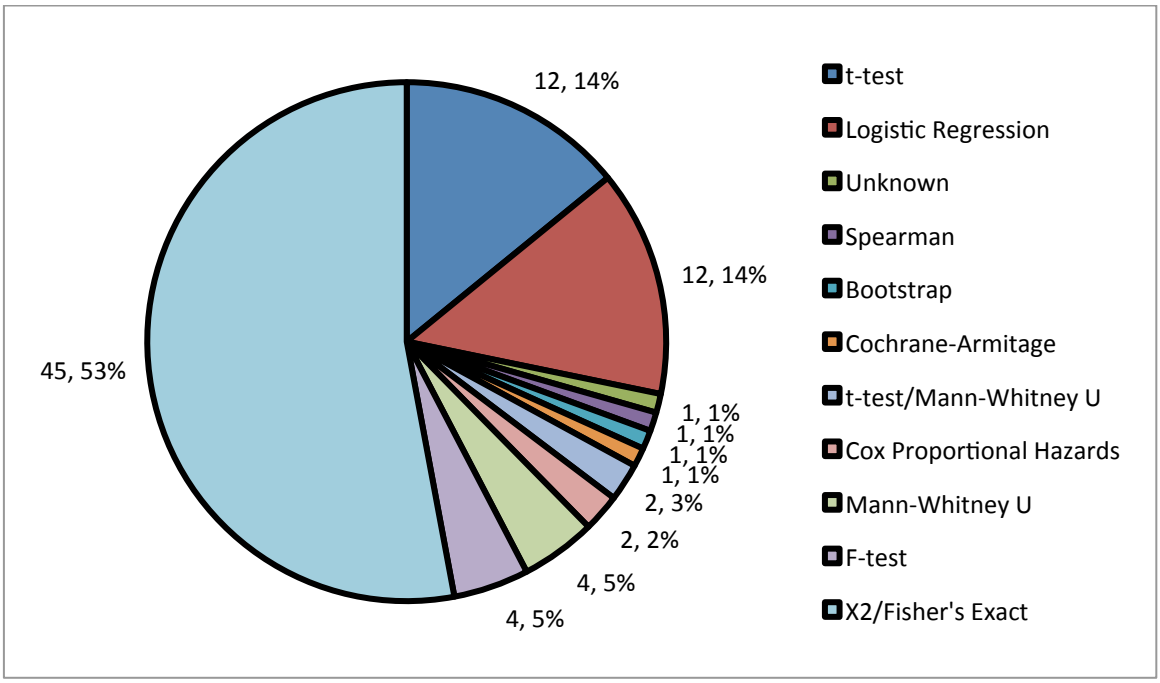


Figure 4.4: Pie chart of the distribution of statistical tests for CLNM studies. Each unit represents a single statistical test. In one case, the statistical test used to study the relationship of interest was not given by the authors.

4.5 Methods: simulation study

In the previous two sections, the methods and results of the literature review were outlined. The focus of the chapter now shifts to the simulation study, conducted in parallel. We have already established that continuous variable categorization is a poor practice and a potential source of cut-point variability. It is important to quantify the extent to which use of the "optimum p-value" method and multiple testing influences study outcome.

This section uses statistical simulation to assess the influence of dichotomization on the dataset established in Chapter 3. A series of cut-points are applied to a subset of the dataset and the "optimum cut-point" is chosen using the lowest p-value criterion. The resulting cut-point is then tested with a set of novel cases. The intent is to establish the extent to which the cut-point chosen is data dependent. Further, the optimum p-value method is applied to a randomized dataset. The intent is to establish the extent to which multiple testing can inflate the number of results that are deemed statistically significant.

4.5.1 General methodology

A series of 4 statistical simulations were designed using R-project. The overall workflow is diagrammed in Figure 4.5. A sample of 199 patients, representing the database described in Chapter 3, was constructed using the pathological depth, pathological CLNM status, 2-year all-cause survival (ACS). 8 cases were excluded because of missing depth of invasion values, leaving 191 patients. For the purpose of survival analysis, 9 patients were excluded for reasons given in Chapter 5, leaving 182 patients. All statistical tests were deemed significant if p was less than or equal to 0.05.

4.5.2 Simulation 1-A: cervical lymph nodes metastasis

First, a simulation sample was selected randomly from the CLNM dataset. The sample size was n = 50, reflecting the median sample size of the CLNM publications documented during literature review. From the remaining cases a validation set of 50 cases was randomly selected. The relationship between depth of invasion and CLNM was evaluated using binary logistic regression.

A series of cut-points ranging from 1 mm to 20 mm were used to classify the depth of each simulation sample. Depth values less than, or equal to, the cut-point were coded 0 while those greater-than were coded 1. Logistic regression was used to evaluate the relationship between the dichotomized depth and CLNM. The minimum p-value was used to select the ''optimum cut-point''.

The simplified p-value correction formula of Altman et al. with $\varepsilon = 10\%$ was used to calculate an adjusted p-value for the ''optimum cut-point''. The ''optimum cut-point'' was used to dichotomize the validation sample. Logistic regression was used to evaluate the relationship between depth and CLNM. This procedure was iterated 1000 times.

4.5.3 Simulation 1-B: cervical lymph nodes metastasis with randomized outcomes

A simulation was carried out, as described in simulation 1-A, with one alteration. In this simulation, the order of the CLNM variable was randomized relative to the depth variable. In this way, a dataset was generated in which the relationship between depth and CLNM status was random.

4.5.4 Simulation 2-A and 2-B: survival

The procedure did not differ from that of 1-A and 1-B, with the exception that all-cause survival was the outcome of interest instead of CLNM and the simulation and validation sample sizes were both n=90, reflecting the median sample size of the survival publications documented during literature review. The Cox proportional hazards test was used in place of logistic regression to evaluate relationships between continuous variables and survival. The log rank test was used to evaluate the relationship between dichotomized variables and survival.

4.5.5 Section summary

In this section the structure of a series of four statistical simulations was outlined. The purpose of these simulations is to quantify the degree of cut-point variability and p-value inflation that results from continuous variable categorization and multiple testing.

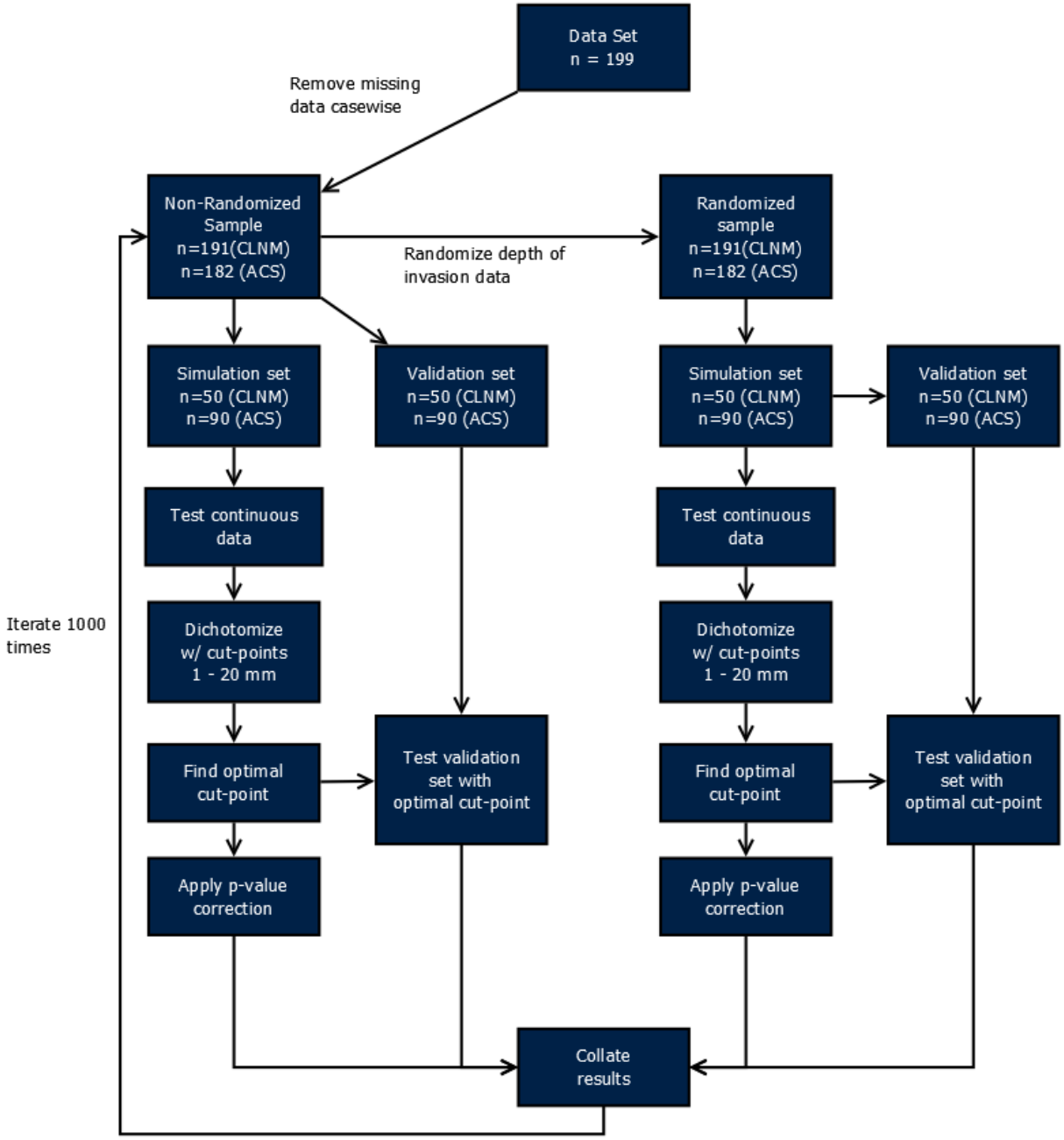


Figure 4.5: Workflow for simulation study.

4.6 Results: simulation study

In the previous section we established the procedure for a series of four statistical simulations meant to clarify the potential impact of cut-point selection on the oral cancer depth of invasion literature. This section presents the results of these four simulations. Cervical lymph node metastasis and survival are presented separately and randomized and non-randomized samples are compared for each of the two outcomes. The majority of the results are presented graphically.

4.6.1 Cervical lymph node metastasis

The results of simulation 1 are shown in Figure 4.6 and Figure 4.7, on the following three pages.

4.6.2 Survival

The results of simulation 2 are shown in Figure 4.6 and Figure 4.8, on the following three pages.

4.7 General discussion

The previous four sections presented the methods and results of the literature review and simulation studies. The general discussion brings these two parallel works together and identifies common themes. The implications of the combined results to the interpretation of the depth of invasion literature are considered.

In the literature review we established that cut-point variability is a widespread problem in the depth of invasion literature for oral cavity cancer. The sample characteristics vary widely from study to study and highlight one source of variation. Further, many studies fail to clearly describe their statistical methodology and reason for choosing a putative cut-point. The literature is suspicious of the optimum p-value method of cut-point selection.

The simulation study established the potential effect of widespread use of the optimum p-value method. Using such a practice leads to cut-point values that are highly data-dependent and is a likely source of variability. As well, dichotomization leads to artificial lowering of p-values, potentially leading to falsely significant results in up to 35% of cases where no such relationship exists.

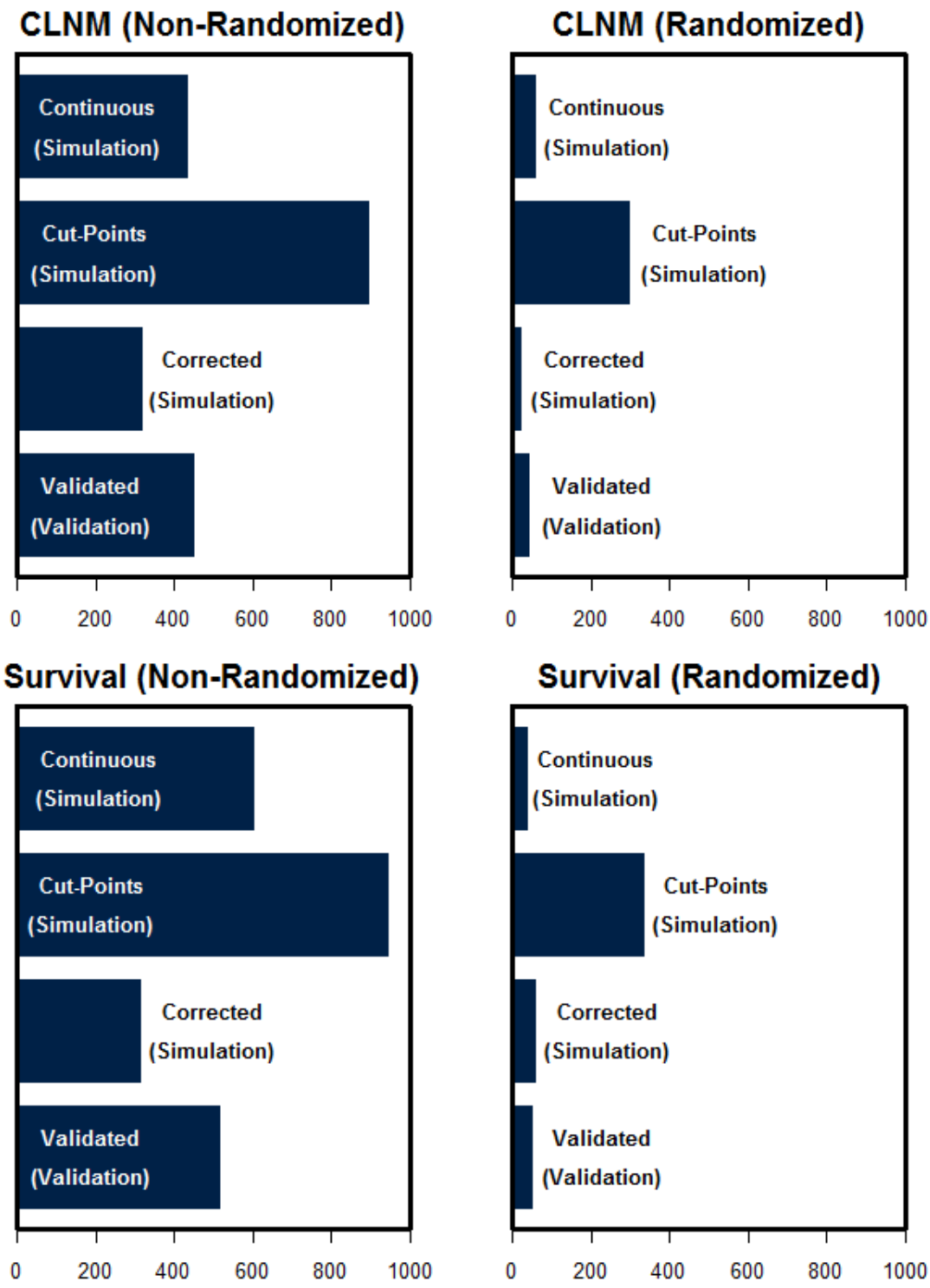


Figure 4.6: Results of simulation for CLNM and survival outcomes. Each x-axis unit represents a single iteration. The blue bars represent the number of significant tests. In the CLNM simulations, simulation and sample sets of n=50 were used while in the survival simulations, simulation and validation sets of n=90 were used. In all cases, the dichotomization overestimates the relationship between depth of invasion and outcome. In the randomized samples, significant relationship were discovered in 30-35% of cases, compared to the expected value of 5%.

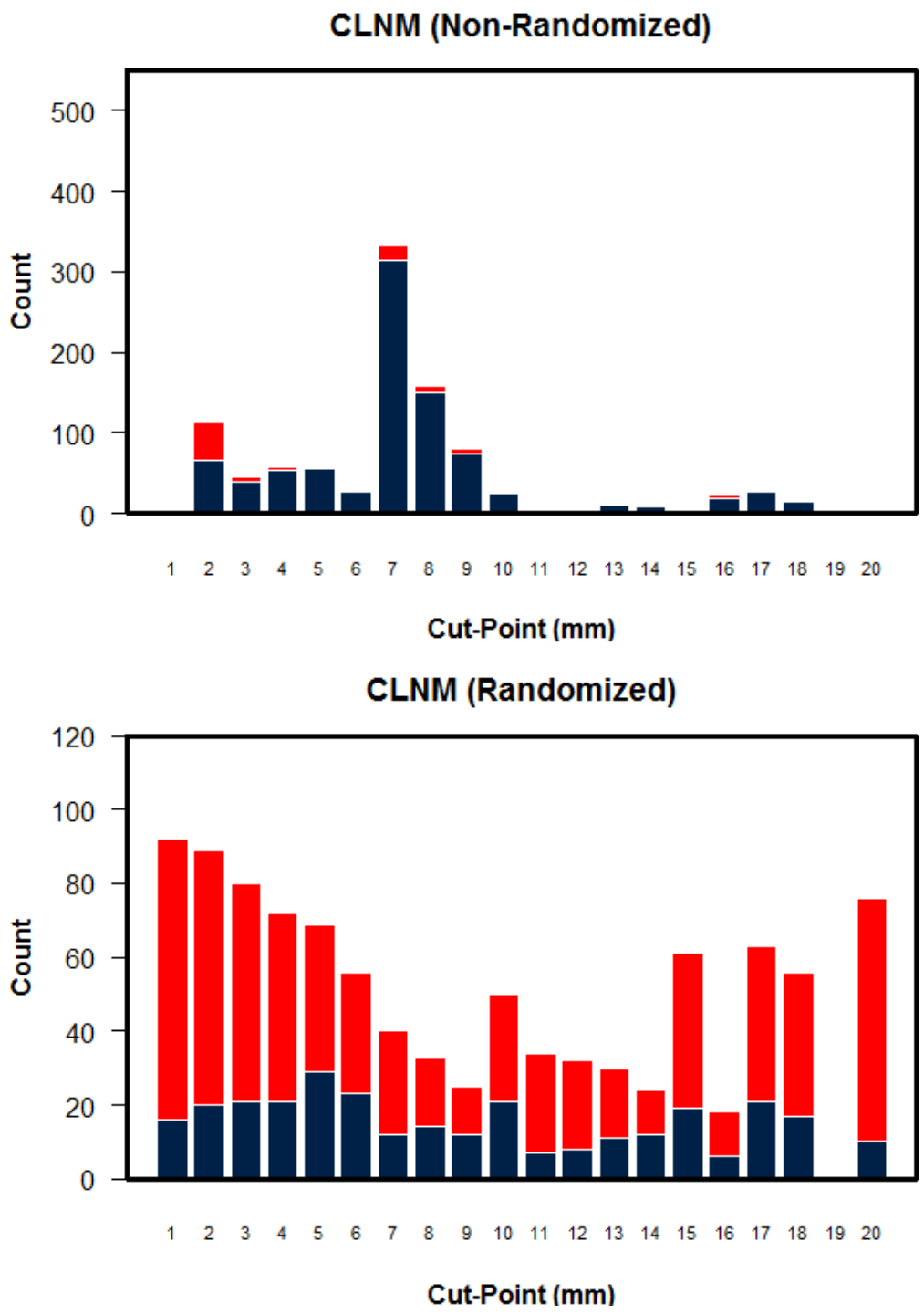


Figure 4.7: "Optimum" cut-point values for CLNM. "Count" represents the number of iterations in which the result was found. Significant results are shown in blue while non-significant results are shown in red. In the non-randomized simulation, the majority of optimal cut-point values were significant and concentrated at 7 mm. However, in the non-randomized simulation, far fewer results were significant and the cut-point value was distributed amongst all possible values.

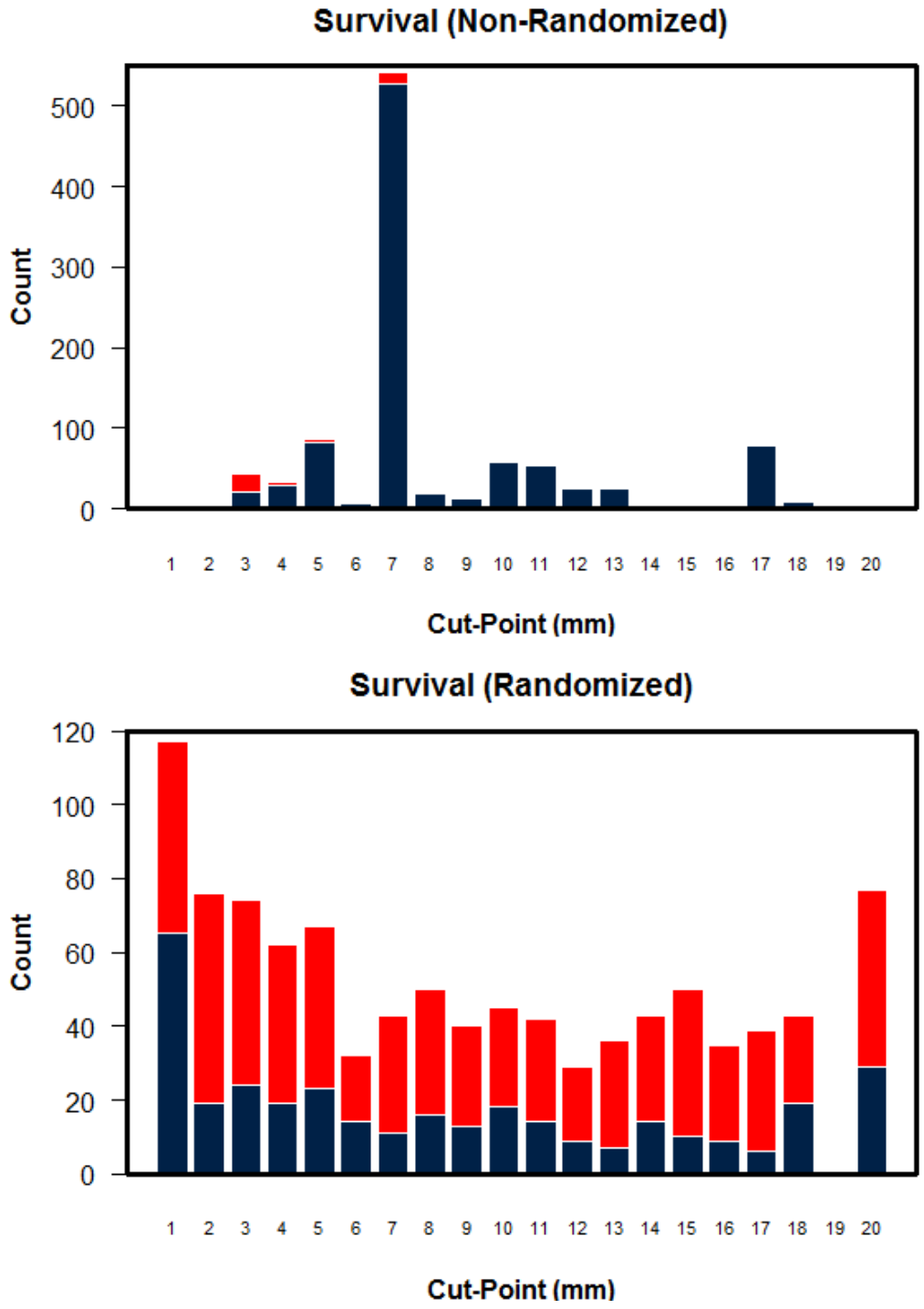


Figure 4.8: "Optimum" cut-point values for survival. "Count" represents the number of iterations in which the result was found. Significant results are shown in blue while non-significant results are shown in red. In the non-randomized simulation, the majority of optimal cut-point values were significant and concentrated at 7 mm. However, in the non-randomized simulation, far fewer results were significant and the cut-point value was distributed amongst all possible values.

The implications of the results are not trivial. If the "worst case scenario" is considered, in which the majority of depth of invasion studies are based on optimum p-value, the current body of literature must be reconsidered.

4.7.1 Variable sample selection

There was widespread heterogeneity in the methodology of the evaluated publications. As noted by the Pentenero review [193], despite the large number of publications, any meta-analysis is impossible.

Thickness and depth are fundamentally different in definition and therefore in measurement, yet many publications failed to provide methodology. This is a crucial element of study design as the two entities are not interchangeable. In many cases with clear methodology, thickness or depth were used in error, contributing to the perception of interchangeability and confusing the reader.

Given this less than rigorous treatment of nomenclature, it is difficult to differentiate thickness and depth for the purposes of this literature review. No further attempt will be made to do so. For the remainder of the discussion, depth is used to refer to both thickness and depth as a single entity.

4.7.2 Does a relationship exist?

The strongest body of evidence exists for a relationship between tumour depth and CLNM. CLNM was evaluated in 51 studies reporting 85 statistical tests on 59 samples. This is compared to survival, with 25 publications reporting 52 statistical tests on 30 samples. However, the relationship between depth and these outcomes is by no means straightforward.

Any conclusions about specific relationships are skewed by widespread sample heterogeneity. While the publications share a common goal, dissimilarities in approach to sample inclusion and exclusion criteria limit cohesion. The result is a series of publications evaluating depth in a variety of specific situations, rather than a progressive growth of evidence.

That said, it is reasonable to conclude that depth can be used as an indicator of CLNM and survival.

Of statistical tests treating depth as a continuous variable, the ratio of significant tests to non-significant tests was 3:2 for both CLNM and survival. When depth was categorized, the resulting ratios are 29:8 and 35:11.

However, it is likely that the evidence for depth is over-represented by the literature. If the problem of multiple testing is considered, significant results using categorized depth values are called into question. Indeed, when the Altman p-value correction was applied, the ratio of significant to non-significant tests fell to 20:10 and 20:15 for CLNM and survival respectively. This is compounded by negative publication bias in the literature, which inflates the significant to non-significant ratios.

Disease-related survival is underrepresented relative to all-cause and disease-free survival. In fact, it is difficult to conclude that depth predicts DRS. Only one of two DRS studies using continuous data was significant, though the non-significant study had a sample size of only 26 patients compared to 85 for the significant result. 2 of 10 statistical tests using cut-point values were non-significant. A further 3 were classified as non-significant when the p-value correction was applied for a total of 5 of 10 results. In fact, ACS and DRS had more non-significant than significant results when the p-value correction was applied. Caution is advisable when interpreting the application of the p-value correction formula. However, caution is equally advisable when interpreting uncorrected results using categorized data.

4.7.3 Does a single cut-point value exist?

The range of recommended cut-points found in the literature for both CLNM and survival (1.5-16 mm) reveals there is no clear-cut definitive value. Using CLNM as an example, even if we were to recommend the most frequently chosen cut-point, 5 mm, such a value would only be supported by 50% of the literature.

This range can be partially explained by the variability in inclusion and exclusion criteria. For example, samples including only stage I and II tumours are more likely to offer smaller cut-points than studies including all stages. Other authors suggest the source may be variability in measurement

methodology. The lack of discrimination between depth and thickness supports this theory.

It may be that multiple testing and optimal p-value technique play a major or even dominant role. SPF was thought to offer useful prognostic information in breast cancer. However, like depth in the oral cancer literature, there was disagreement about a nominal cut-point with values ranging from 2.6% to 15.4%. Altman et al. point to multiple testing as a contributing factor and in their commentary remark that:

> *it should be kept in mind that the cut-point obtained [by the optimal p method] is highly data dependent and so would be expected to vary markedly between samples. [197]*

It is likely that the same phenomenon affects the oral cancer literature. This is supported by the simulation study presented by this work. When CLNM and non-randomized depth were used as simulation parameters (1-A), a significant cut-point was found almost 90% of the time. The ''optimum'' cut-point assumed values across from 1-20 mm. Only 45% of the time were these optimum cut-points validated using an independent test sample. Similar results were seen in the non-randomized depth versus survival simulation (2-A) in which 52% of cut-points were validated.

In both the CLNM and survival simulations (1-A and 2-A), one cut-point emerged as the most frequent optimum value. It is tempting to conclude that this dominant value is the ''best'' cut-point. However, even the most frequent cut-point never occurred more than 50% of the time. At best, a crude confidence interval for the cut-point can be established. In the review of the literature, 30% of the statistical tests used sample sizes of less than 50, indicating that cut-point variability may be high.

A typical graph of p-value as a function of cut-point from the simulation is shown in Figure 4.9. It was not uncommon for multiple cut-points to have p-values less than 0.05. In many cases, the p-value of the optimal cut-point differed only slightly from that of adjacent values. Similarly, it was not uncommon for these graphs to show multiple ''local minima'' with p-values less than 0.05. It is obvious that slight differences in p-value are arbitrary and, thus, the choice of ''optimum'' cut-point is dependent on small variations in the data sample.

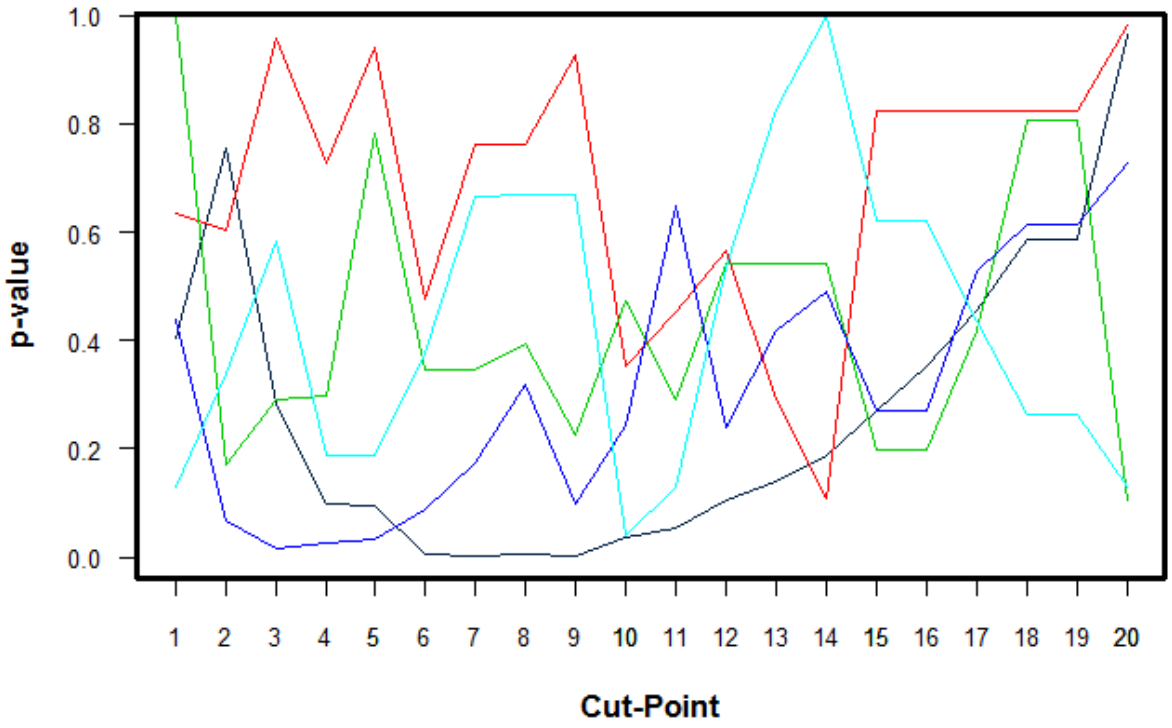


Figure 4.9: Example of cut-point significance variability. Each line represents a single randomly chosen simulation set, tested after being dichotomized at cut-points from 1 - 20 mm.

These results suggest that cut-point selection is data-driven. Assuming that the use of optimal p-value testing is widespread amongst authors not reporting cut-point selection methods, it reasonable to conclude that this is a major source of cut-point variability in the oral cancer literature.

4.7.4 What role does multiple testing play?

One of the purposes of this literature review was to evaluate the extent of multiple testing in the OCSCC literature. Only a small number of publications stated that multiple cut-points were evaluated and the optimal value chosen based on the minimum p-value. As an example, Liao et al. tested cut-point values ranging from 5-17 mm and p-values ranging from <0.001 (9-12 mm) to 0.238 (17 mm).

The number of publications that stated their reasons for choosing their cut-point was also limited. Some studies chose the median thickness value as the cut-point. Others chose a cut-point based on previous publications in the OCSCC literature. There were several cases in which authors commented that a previously published cut-point produced a non-significant result, then proceeded to give their own novel significant cut-point, implying multiple testing.

Some researchers chose a cut-point value based on the sensitivity and specificity of the results of their study. This is more common in studies of CLNM because any decision rule with regard to neck dissection must minimize false-negative results. In practice, this means that researchers tend to choose the smallest significant cut-point value. This is a variant of the optimum p-value that still employs multiple testing, but to a lesser extent.

When methodology is not given, it seems reasonable to conclude that multiple testing method has been used to find a significant result. This is a direct result of negative publication bias in the literature. Once depth and outcome data have been collected, searching for and reporting a ''highly significant'' cut-point greatly improves that author's chance of successful publication.

4.7.5 Multiple testing and false negative results

One problem associated with multiple testing was highlighted by the simulation studies using

randomized data (1-B, 2-B). Theoretically, when a statistical test specifies an p-value threshold of 0.05, the incidence of falsely-positive results should be no more than 5%. We expect that only about 5% of the simulation results should be significant purely by chance.

In practice, using the optimal p-value method, a staggering 30% of CLNM and 34% of survival results were significant. As expected, the optimal p-value method is over-optimistic and akin to using an p-value threshold of 0.30 or 0.34. Thus, when authors use an optimal p-value method for selecting a cut-point, there is a large probability of spuriously significant results. This is a contributor to cut-point variability and also indicates that readers must be cautious when interpreting the depth literature.

4.7.6 Addressing cut-point variability and multiple testing

The p-value correction formula proposed by Altman at al. represents a method for addressing the over-optimism evident in the randomized depth simulations [197]. Ideally, the corrective factor would reduce the incidence of false-positive results to 5% from 40% plus. Assuming ε = 10%, solving the formula shows that only results with uncorrected p-values less than 0.002 should be considered significant. When applied to the randomized CLNM data, the correction was conservative with 2.5% of tests significant. Conversely, when applied to the randomized survival data, the correction factor was slightly optimistic with 6.1% of tests significant.

There are a few potential explanations for this discrepancy. First, the corrective formula assumes that the continuous explanatory variable is normally distributed. In this work, depth measurements were skewed to the right and undoubtedly affects the outcome when the corrective factor is applied. It is unclear why CLNM produces a conservative correction while survival produces the opposite.

The corrective formula assumes that the value of ε = 10%. That is, the smallest and largest 10% of the range of the continuous variable are discarded. The rightward skew of the oral cancer depth data produces a situation in which 'optimal cut-points' cluster at small thickness values. Discarding the smallest 10% of potential cut-points would remove a number of candidates previously reported as significant in the literature. As such, only larger thickness values (>20 mm) were discarded.

The correction formula is a crude method of reducing spuriously significant p-values. That is not to say it is not of value. In any situation, it is more desirable than the uncorrected alternative. Perhaps the best approach to evaluating the significance of tumour depth is to avoid classification in the first place. This is recommended by Altman et al. and avoids the loss of statistical power associated with dichotomization.

4.8 Conclusions

This chapter has produced evidence that much of the oral cancer literature that deals with the depth of tumour invasion may be based on poor statistical methodology, selecting cut-points based on the "optimal p-value". At the very least, many authors do not explicitly define the methods with which reported cut-points were derived.

Further, the simulation study confirms that the multiple testing implied by the use of the "lowest p-value" method results in cut-point value selection that is highly sample dependent, and a likely source of cut-point variability seen throughout the oral cancer depth literature. In fact, use of the "optimal p-value" method results in a 6-7 fold increase in false positive statistical tests, and is akin to accepting a p-value of less than 0.35 as significant. Given that the use of the "optimal p-value" method may be wide spread, there is good reason to question the available literature.

The dichotomization of continuous data is a poor statistical technique that, at the very least, results in the loss of valuable information. Ideally, continuous data should be evaluated using appropriate statistical techniques, such as logistic regression and Cox proportional hazards, to evaluate relationships with categorical outcomes and survival.

In many cases a cut-point value might be desirable. For example, in some cases continuous data might have a non-normal distribution that is not amenable to any statistical transformation. In these situations, analysis might not be possible without categorization. Alternatively, medical decision making often requires discrete boundaries for classifying patients into low and high-risk groups in a

manner that is standardized across institutions.

However, when categorization is necessary, the "optimum p-value" method must be avoided. Cut-points should be selected prior to data analysis, based on previous literature or the median value, to avoid problems with multiple selection. The use of three or more cut-points results in less loss of information as such groupings more closely approximate continuous distributions.

Much of the variability associated with choosing cut-points is unavoidable. As such, the search for a "best" cut-point is likely to result in nothing more than further argument. At some level, the choice is arbitrary and, acknowledged as such, the only way forward is to choose a value and retain it.

5 Data summary and univariate analysis

5.1 Introduction

The patient database described in Chapter 3 contains a wealth of information. However, that information is worthless unless it can be presented and summarized in a coherent manner, and used to further our understanding of oral cancer. The goal of this chapter is to bring together the content of the database in a manner that methodically outlines each factor, its contribution to occult cervical lymph node metastasis (OCLNM) prediction and its prognostic value. The downfall of this methodical approach is repetition. However, the information is compartmentalized in such a way that the results and discussion for each factor are presented together, as a single unit. The result is a library of predictive and prognostic factors, evaluated on a relatively large patient sample, that are of key importance in current oral oncology practice.

The first step in any multi-factor model is to identify factors that are independent predictors of the outcome of interest. In the realm of oral cavity cancer a large number of factors are known to contribute to the ability to predict all-cause, disease-related and disease-free survival. A similarly large number of factors are known to be significant predictors of cervical lymph node metastasis. These factors can be either patient or tumour related.

This chapter aims to identify those factors available on retrospective chart review and image analysis, which are independent predictors of the outcomes of interest: all-cause survival (ACS), disease-related survival (DRS), disease-free survival (DFS) and OCLNM. An important element of this analysis is to compare the univariate results with the work of other authors to identify and evaluate any disagreement with established trends. Those factors identified as significant will be carried forward and used to build multivariate predictive models in Chapter 6.

5.1.1 Chapter goals

The goals of this chapter are:

1. *To present methods and procedures that are unique to univariate analysis of the data.*

2. *To summarize surgical data, occult cervical lymph node metastasis data and survival data.*

3. *To present the results of the univariate analysis of patient and tumour related predictive and prognostic factors.*

4. *To discuss the results of univariate analysis and compare these results with previous work in the oral cancer prognosis.*

5. *To determine which prognostic and predictive factors are suitable for inclusion in multivariate models of OCLNM and survival, to be constructed in Chapter 6.*

5.1.2 Organization of sections

The methodology section provides the reader with precise definitions of the outcomes of interest. Methodology for each prognostic and predictive factor is given in Chapter 3. The next three sections summarize the data on surgical procedures and treatment received, occult cervical lymph node metastasis and the three subtypes of survival, giving the reader an overview of the patient sample.

The remaining sections deal with each prognostic and predictive factor individually. Here, the thesis departs from traditional formatting. The factors are compartmentalized according to type: patient, clinical, staging, pathology and measurement, each comprising a section. Where possible, data is presented in graphical or tabular format at the end of each section. Appendix IV provides detailed information about data summary, OCLNM and survival results.

For each individual prognostic and predictive factor, the results and discussion take place as a single unit, rather than in separate sections. In doing so, it is hoped that some element of cohesiveness can be maintained for each factor. Finally, a concluding section summarizes the content of the chapter.

5.2 Methodology

The following section outlines methodology specific to this chapter. Occult cervical lymph node metastasis is defined. The criteria for 2-year all-cause, disease-related and disease-free survival are given. Statistical methods are mentioned, though a more detailed background in found in Chapter 6.

5.2.1 Definition of occult metastasis

Occult cervical lymph node metastasis is defined as the presence, on pathological examination, of metastatic SCC in a patient classified as clinically N0 following clinical examination and staging MRI.

The presence of CLNM on at the time of primary tumour resection was defined in two ways:

> *a) The pathologist confirmed presence of SCC in lymph nodes sampled by neck dissection at the time of surgery.*
>
> *b) The pathologist confirmed presence of SCC in cervical lymph nodes sampled within 2 months of primary surgery, in patients that had not undergone adjuvant radiotherapy or chemotherapy in that time period.*

5.2.2 Definition of 2-year ACS, DRS and DFS

5.2.2.1 Two year survival period

Two years was chosen as the survival period. Due to the retrospective study design, the follow-up period was variable, despite the 5-year follow-up plan adopted as standard by both the JRH and UCLH. Two years was appropriate since the majority of patients had in excess of two years of follow-up. In addition, the majority of disease-related deaths and recurrences occurred within the first 24 months.

All time periods were measured from the date of primary surgical resection to the date of the event or censoring. All events occurring after the two-year follow-up were censored. Patient follow-up was recorded to a maximum of five years.

5.2.2.2 All-cause survival (ACS)

An all-cause survival (ACS) event was defined as any patient death within the two-year follow-up period, with the exception of peri-operative deaths related to surgical complications. Patients were censored if they were lost to follow-up or if they were still alive at their last documented follow-up. A summary of event and censored data is shown in Table 5.1.

5.2.2.3 Disease-related survival (DRS)

A disease-related survival (DRS) event was defined as any patient death in the two-year follow-up period confirmed by chart review to be a result of their OCSCC, with the exception of peri-operative deaths related to surgical complications. Patients were censored if they were lost to follow-up, if they were still alive at their last documented follow-up or died in the follow-up period as a result of pathology unrelated to OCSCC. A summary of event and censored data is shown in Table 5.1.

5.2.2.4 Disease-free survival (DFS)

A disease-free survival (DRS) event was defined as any pathologist-confirmed local, regional or distant disease recurrence or any pathologist-confirmed second primary SCC of the upper aerodigestive tract in the two-year follow-up period. Patients were censored if they were lost to follow-up or if they were free of recurrences or second primaries at their last documented follow-up. A summary of event and censored data is shown in Table 5.1.

Table 5.1: Criteria for all-cause, disease-related and disease-free survival.

	Event	Censored
ACS	Any death	Alive at last follow-up
		Alive at 24 months
		Lost to follow-up
		Event > 24 months
DRS	Disease-related death	Alive at last follow-up
		Alive at 24 months
		Death unrelated to disease
		Lost to follow-up
		Event > 24 months
DFS	Local recurrence	Recurrence-free at last follow-up
	Regional recurrence	Recurrence-free at 24 months
	Distant recurrence	Death
	Second primary tumour	Lost to follow-up
		Event > 24 months

5.2.3 Statistical analysis

Univariate analysis was carried out using SPSS. Where continuous variables were not normally distributed, a transformation was carried out using the formula $\log_{10}(X+1)$ where X is the continuous variable of interest. All categorical variables were dummy coded.

The ability of continuous and categorical variables to predict OCLNM was tested using logistic regression. Survival was tested using Cox proportional hazards. Correlation between ordinal variables was evaluated using the Spearman correlation test.

For all tests, a p-value less than or equal to 0.05 was used to indicate significance. A p-value between 0.05 and 0.10 was used to indicate a trend towards significance. The theory of each statistical test is described in detail in Chapter 6.

5.3 Summary of surgical factors, survival and OCLNM

It is important to characterize the study population by outlining the surgical treatment received. A summary of OCLNM and survival is presented as a prelude to detailed analysis of individual patient and tumour factors in forthcoming sections.

5.3.1 Surgical factors

Surgical approach is an important determinant of treatment outcome. Surgical factors, including approach to the mandible, cervical lymph nodes, reconstruction and adjuvant therapy are summarized graphically in Figure 5.1 and in tabular format in Appendix IV.

5.3.2 Summary of OCLNM

Of 135 patients initially classified as cN0, 25 (18.7%) were found to have cervical lymph node metastases not seen on clinical examination or MRI imaging. These metastases were classified as occult.

5.3.3 Summary of survival data

5.3.3.1 Survival specific exclusions

8 (4.2%) patients were excluded from the survival analysis because they died in the peri-operative period of complications related to their surgery. 5 patients died as a result of post-operative pneumonia. 1 patient died of sepsis secondary to an abdominal perforation. Two patients died of tracheostomy related complications. The exact cause of death was unknown in 1 case. 1 additional patient was excluded because he was lost to follow-up only 19 days post-operatively.

5.3.3.2 All-cause, disease-related and disease-free survival

A summary of all-cause and disease-related deaths is found in Appendix IV. The anatomic locations of disease recurrence and second primaries are also summarized in Appendix IV.

In total, there were 33 ACS events, 29 DRS events and 44 DFS events. ACS, DRS and DFS for the survival sample are plotted in Figure 5.2, Figure 5.3 and Figure 5.4.

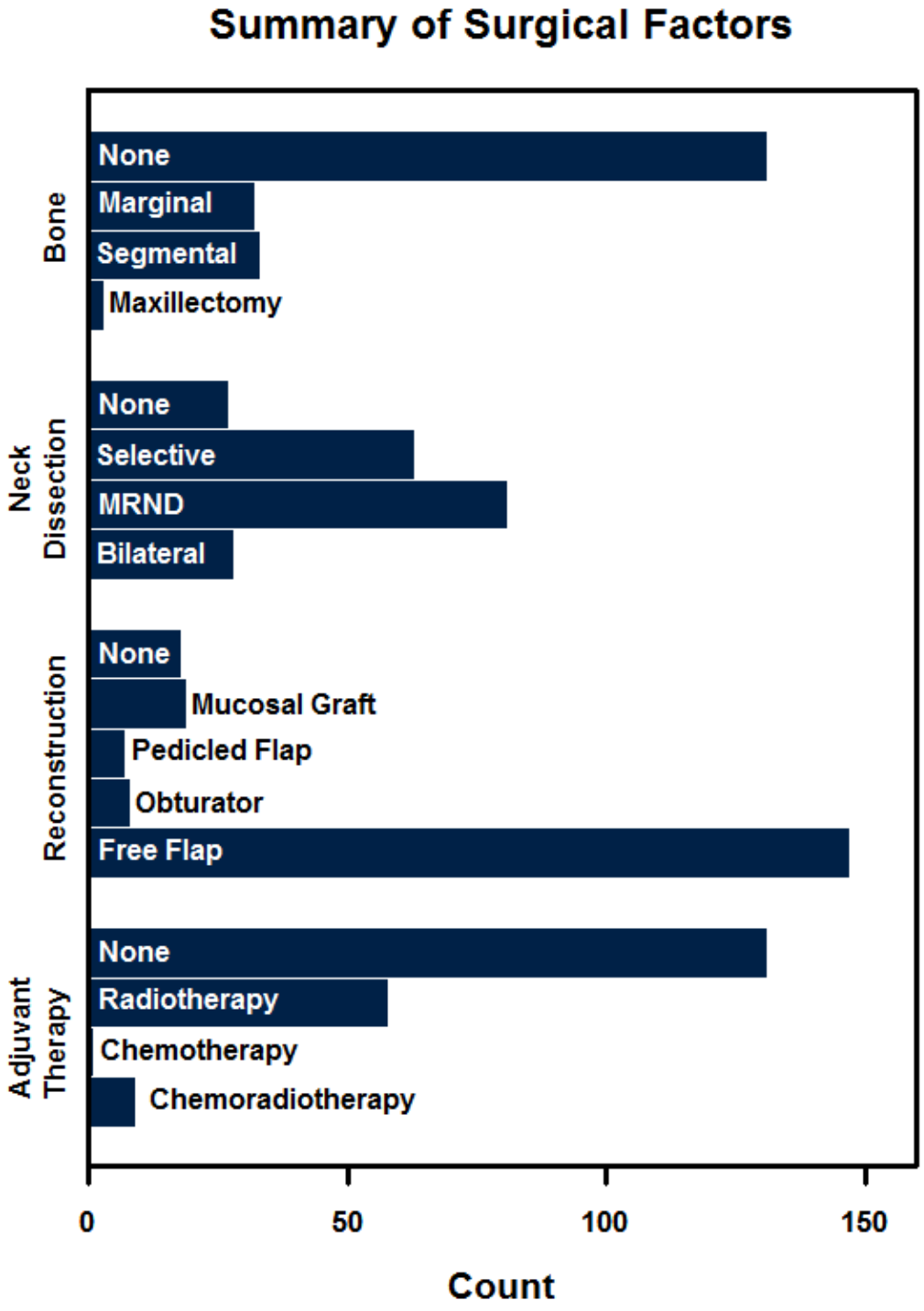


Figure 5.1: Distribution of surgical factors.

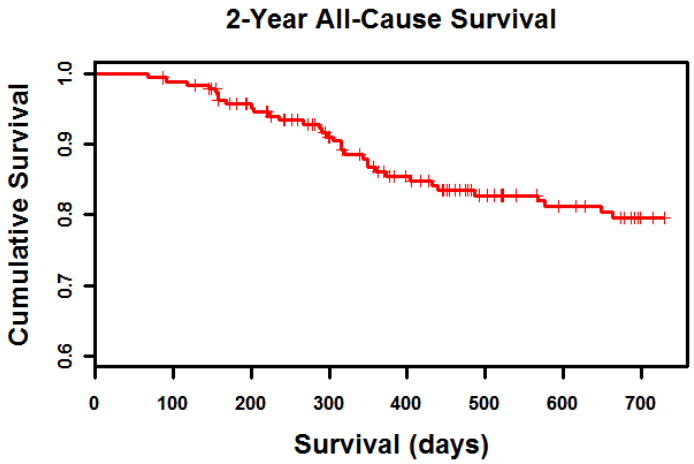


Figure 5.2: Kaplan-Meier survival plot depicting 2-year all-cause survival.

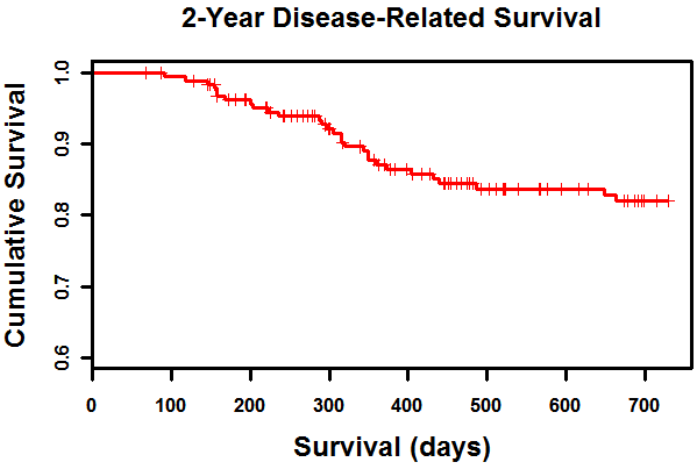


Figure 5.3: Kaplan-Meier survival plot depicting 2-year disease-related survival.

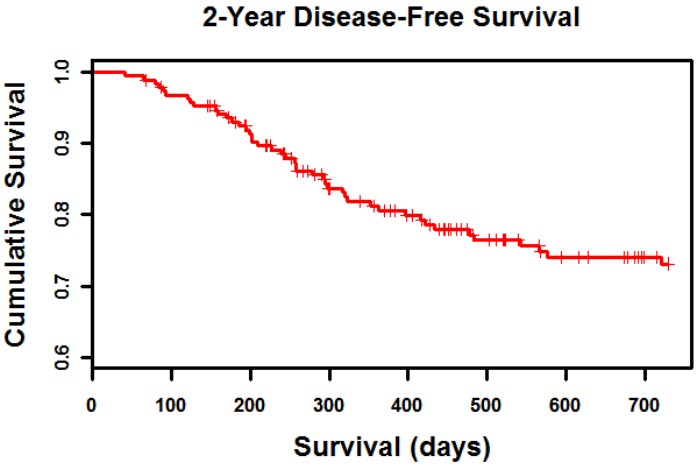


Figure 5.4: Kaplan-Meier survival plot depicting 2-year disease-specific survival.

5.4 Patient factors

This section is the first of five concerning the patient and tumour factors identified for the prediction of OCLNM and survival. For each factor, results and discussion are presented together. Patient factors include age, gender, smoking, alcohol consumption, comorbidity and healthcare institution.

Patient factors are summarized in Figure 5.5 and Figure 5.7, while results for OCLNM and survival are given in Figure 5.6 to Figure 5.10, found at the end of this section. Detailed information is available in Appendix IV.

5.4.1 Age

5.4.1.1 Results

Age was not a predictor of OCLNM, DRS or DFS. There was a trend towards significance with ACS.

5.4.1.2 Discussion

The median age at diagnosis, 61.6 years, was similar to reported values of about 60 years [106]. Increasing age displayed a trend towards significantly decreased ACS. The same relationship was not seen with DRS, DFS or OCLNM. This result is consistent with studies that found no relationship with survival or OCNLM [46, 76, 113-114]. The fact that only ACS decreased with advancing age points to the fact that increasing age is associated with a population-wide increase in mortality [205].

One might expect DRS or DFS to decrease with age on the premise that increased co-morbid illness prevents aggressive treatment. However, patients not undergoing surgery with curative intent were excluded. If elderly patients were disproportionately assigned to chemotherapy or radiotherapy, it would not be reflected in this sample.

5.4.2 Gender

5.4.2.1 Results

Gender was not a predictor of OCLNM or survival.

5.4.2.2 Discussion

The gender distribution was evenly divided between males, 51%, and females, 49%. Reports indicate that OCSCC incidence is approximately twice as high in men [134]. However, the ratio has fallen from nearly 50:1 in the 1950s [134]. The present dataset may reflect an extension of this trend, which has been attributed to increasing smoking and alcohol consumption amongst women [134]. Women may eventually surpass men as the most common gender to suffer from OCSCC.

In the present study, no relationship was found between gender and survival or OCLNM. This is consistent with the published OCSCC literature [46, 113-115].

5.4.3 Smoking and alcohol

5.4.3.1 Results

Neither smoking nor alcohol were predictors of OCLNM or survival.

5.4.3.2 Discussion

This is not consistent with reports that smoking and alcohol are related to survival and the occurrence of second primary tumours [122-124].

The sample size in the present study was likely insufficient to detect any significant relationship. Only 4 patients had second primaries. An longer follow-up might increase the number of second primary tumours detected and give insight into the role of smoking and alcohol in DRS and DFS.

The purpose of gathering smoking and alcohol data was two-fold. First, such information is important for comparing the health of the study population with the general population. Second, we were interested if smoking and drinking information could predict survival or OCLNM. In cases where quantitative information was unavailable, the patient was classified based on qualitative data after a comprehensive chart review. As such, the smoking and alcohol information is not based on rigorous quantitative data. That said, qualitative information obtained about smoking and alcohol assumes that the patient is being truthful and may underestimate reality.

5.4.4 ACE27 comorbidity

5.4.4.1 Results

ACE-27 comorbidity did not predict OCLNM or survival. A summary of comorbidity, broken down by affected system, is shown in Table 5.2.

Table 5.2: Summary of ACE27 comorbidity grouped by affected system.

Category	*None*	*Mild*	*Moderate*	*Severe*
Cardiovascular	*140*	*43*	*14*	*1*
Respiratory	*183*	*13*	*1*	*1*
Gastrointestinal	*186*	*8*	*4*	*0*
Renal	*197*	*1*	*0*	*0*
Endocrine	*186*	*7*	*5*	*0*
Neurological	*196*	*2*	*0*	*0*
Psychiatric	*190*	*6*	*2*	*0*
Immunological	*196*	*1*	*1*	*0*
Malignancy	*178*	*16*	*3*	*1*
Substance Abuse	*174*	*3*	*21*	*0*
Overall	*93*	*59*	*34*	*12*

5.4.4.2 Discussion

While we do not expect comorbidity to affect DRS and DFS, it is surprising that there was no relationship with ACS. Comorbidity is a major source of mortality in the general population; however it may be that the short 2-year follow-up of the study prevented this relationship from being seen. These results are not consistent with studies that show a relationship between comorbidity and survival [126-128]. This work found no relationship with OCLNM. It is not likely that comorbidity reflects the metastatic potential of oral cancer.

The methodology of the ACE-27 index could not be followed precisely using information available in the patient records. Body weight information was not recorded at the time of initial record review and unavailable for comorbidity evaluation. All stated comorbidities do not account for obesity.

In some cases, quantitative data required for comorbidity classification was not uniformly available. Examples include blood pressure and FEV1 measurements. In cases where the disease was

documented but measurements unavailable, "mild" was assigned.

A similar situation existed for categories requiring temporal information. Examples include the time since myocardial infarction and the time since diagnosis and treatment of solid malignancies. Again, the category "mild" was assigned if the required information was unavailable.

In some cases, the requirements for the various categories is subjective. For example, in the "Substance abuse - Alcohol" category "active abuse with social, behavioural or medical complications" is used to indicate moderate comorbidity. One could argue that all individuals with active substance abuse have "behavioural complications".

5.4.5 Institution

5.4.5.1 Results

Institution affected neither OCLNM nor survival.

5.4.5.2 Discussion

This indicates that there was no difference in the standard of care or tumour severity between the JRH and UCLH. It can be inferred that it is suitable to combine cases from these two institutions into a common dataset.

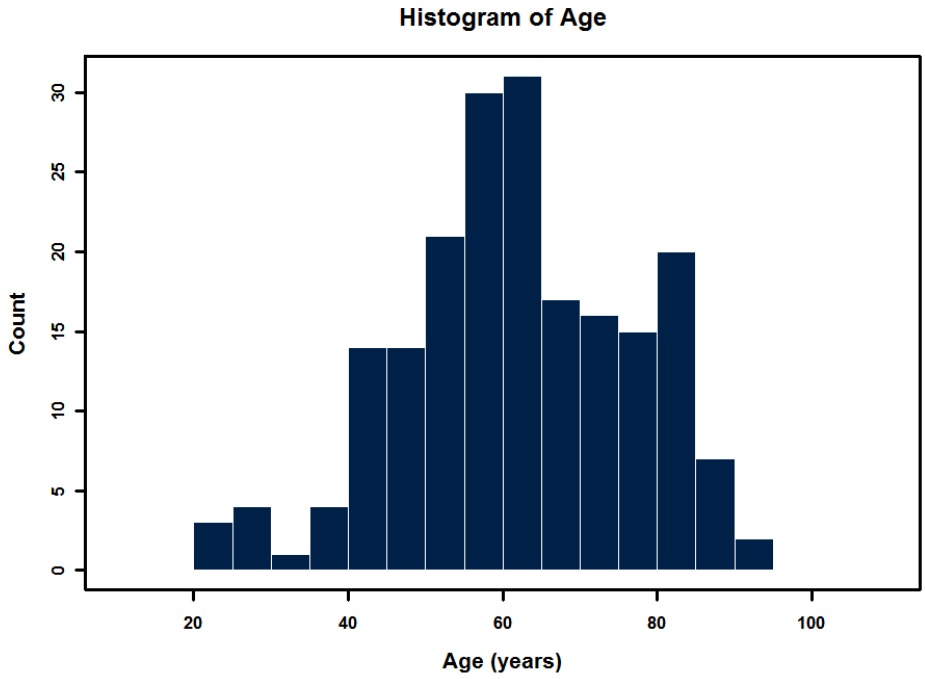


Figure 5.5: Histogram of the distribution of age. Each bar represents a five year time interval.

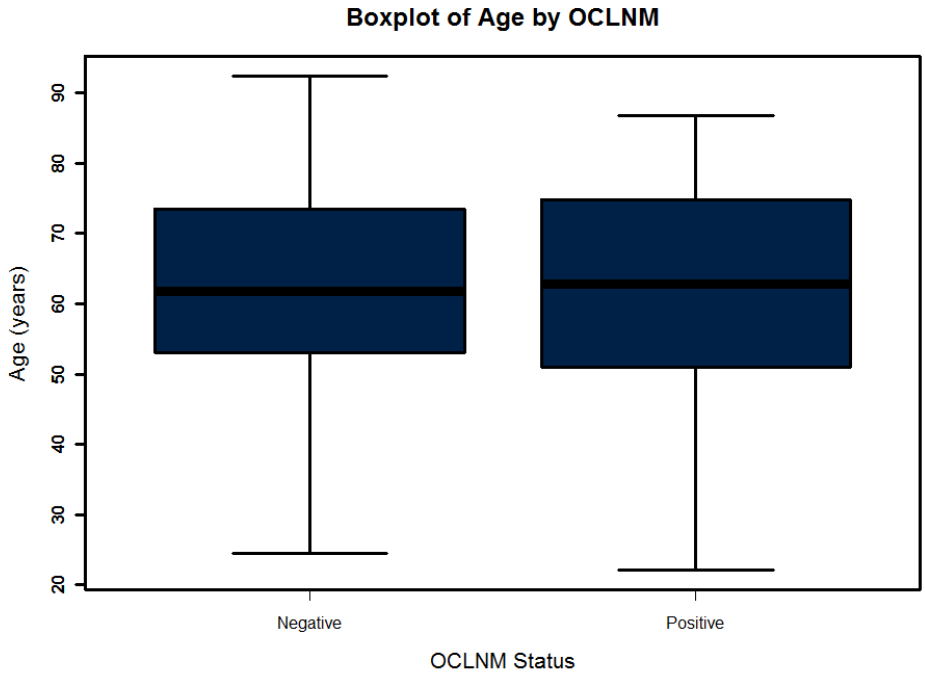


Figure 5.6: Box plot of age by OCLNM. The blue boxes represent the interquartile range (IQR) while the black line within the blue boxes represent the median. The whiskers extend a distance of 1.5 times the IQR from the upper and lower edges of the box.

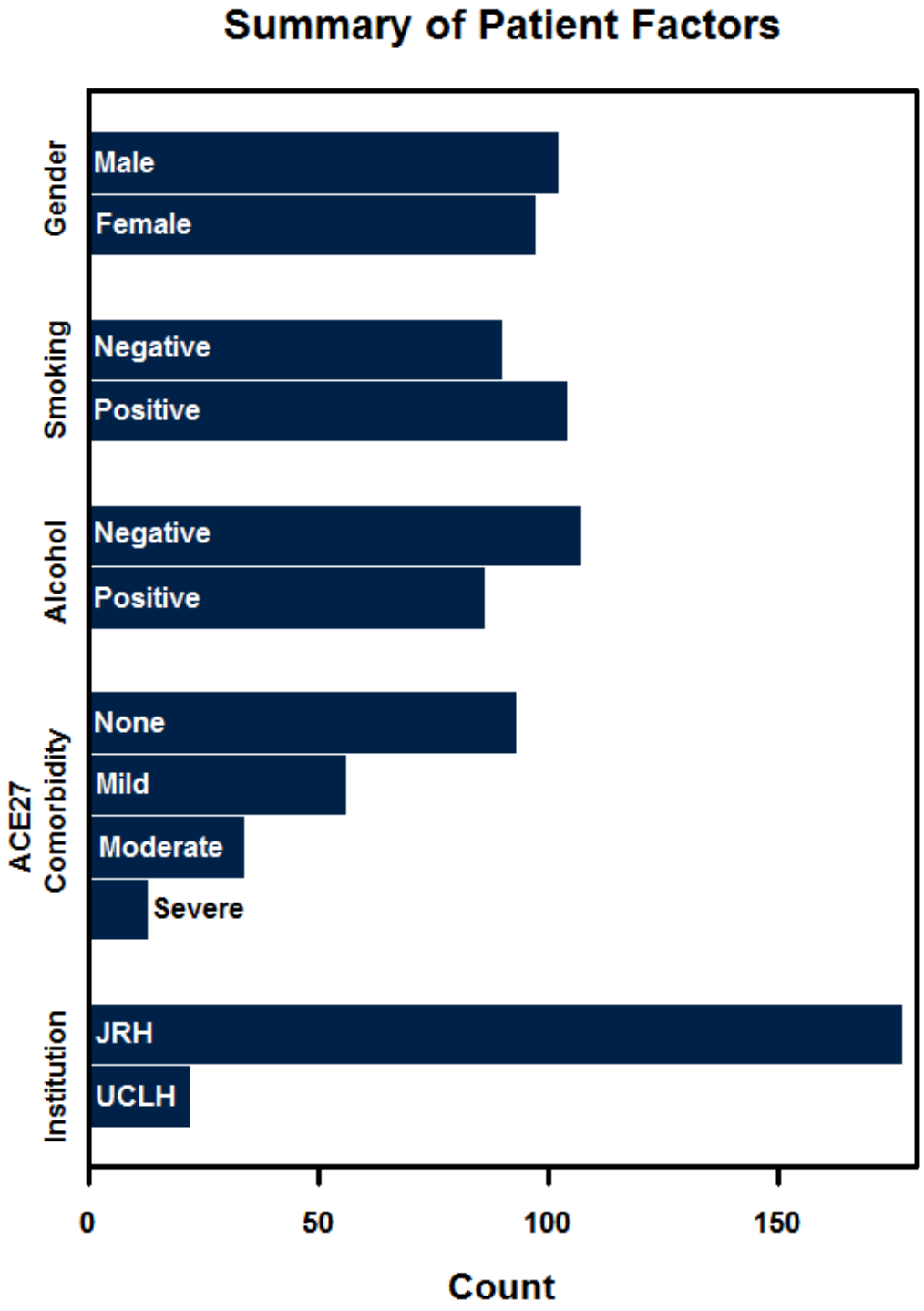


Figure 5.7: Distribution of patient factors. JRH: John Radcliffe Hospital; UCLH: University College London Hospital; ACE: adult comorbidity evaluation

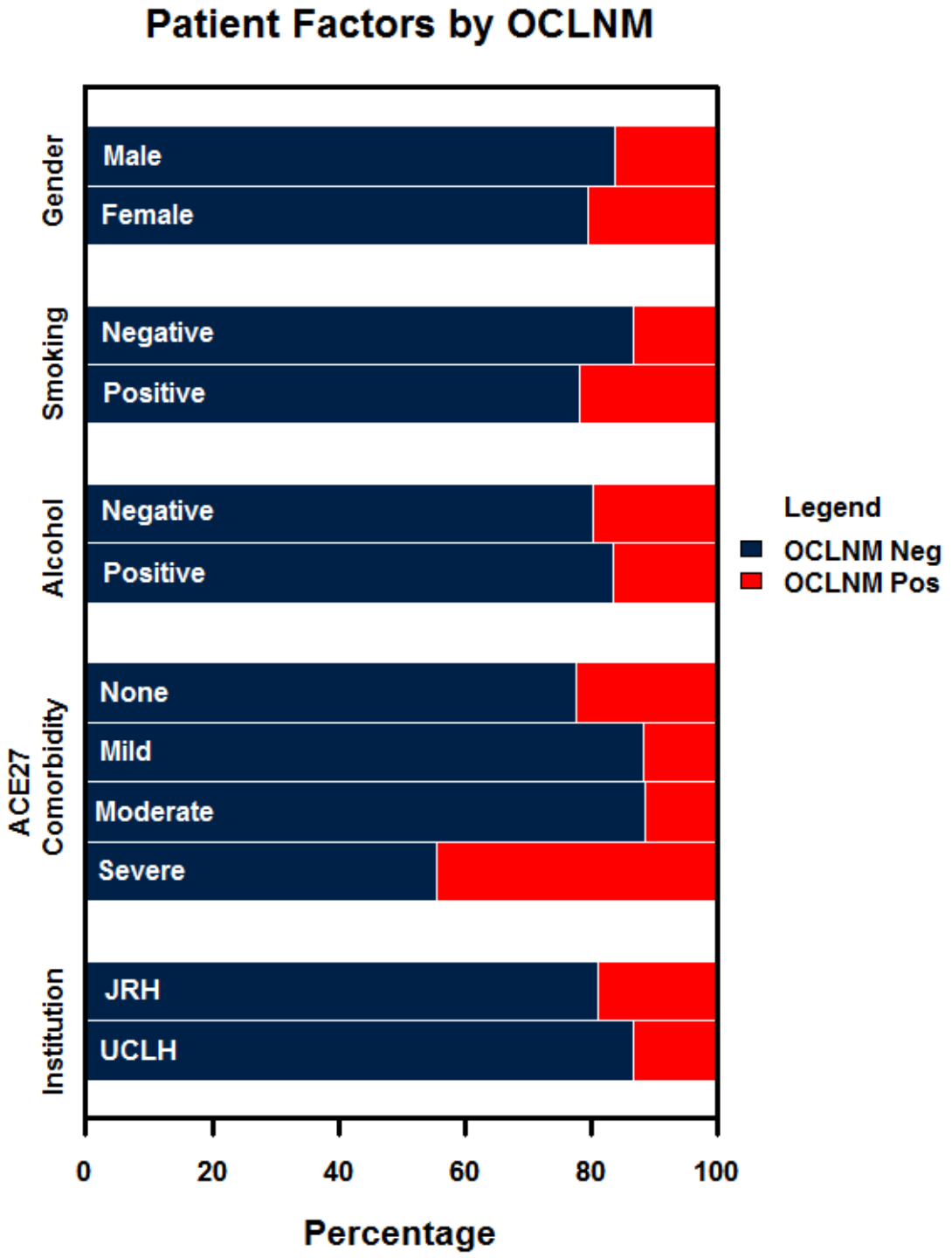


Figure 5.8: Patient factors by OCLNM status. Proportion of cases negative for OCLNM (blue) and positive for OCLNM are shown (red). OCLNM: occult cervical lymph node metastasis; JRH: John Radcliffe Hospital; UCLH: University College London Hospital; ACE: adult comorbidity evaluation.

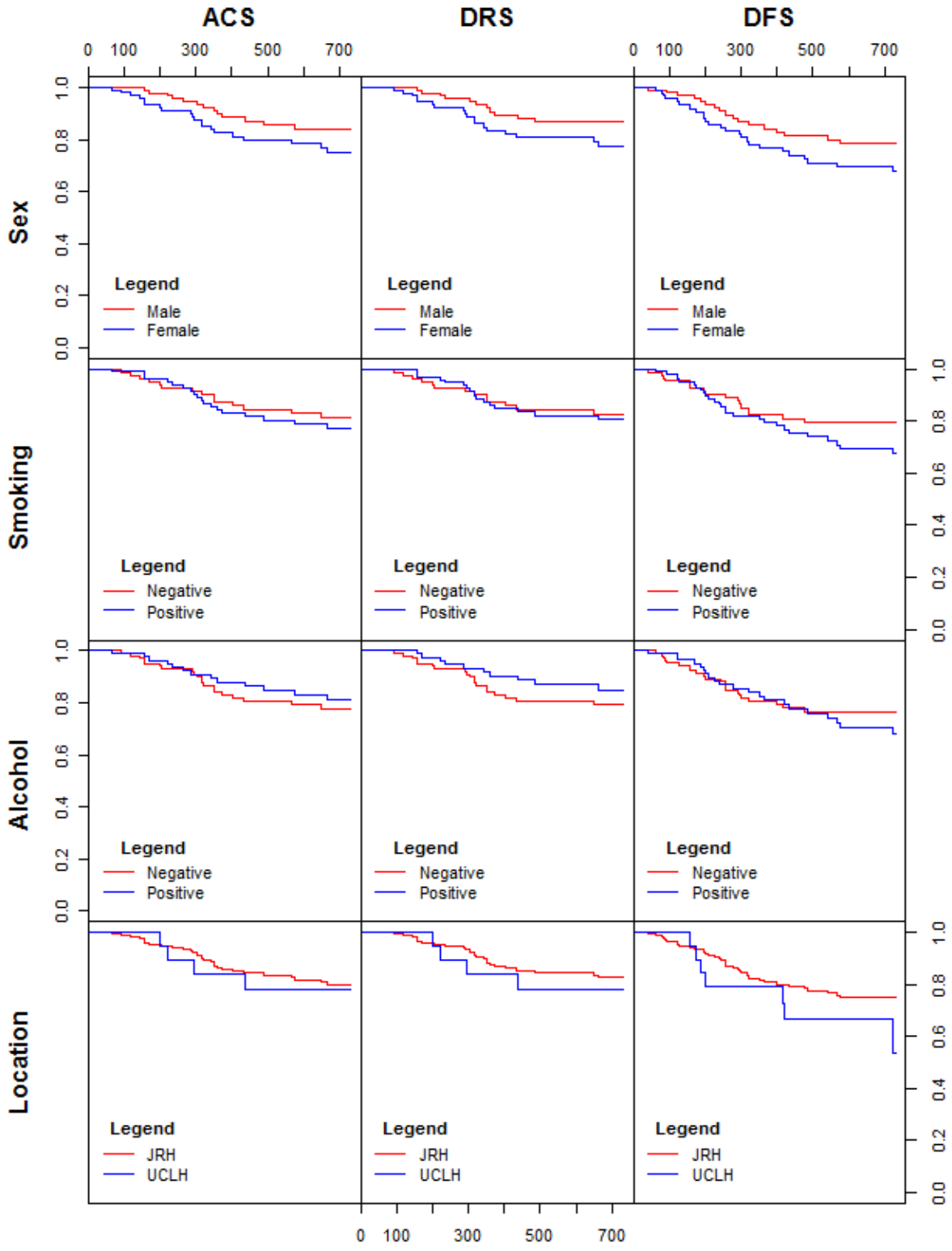


Figure 5.9: Survival by patient factors (1 of 2). Each square represents a Kaplan-Meier. The components involved each survival plot are indicated by the intersection of the patient factors (left margin of grid) and the type of survival (top margin of grid). The x-axis scale of each plot is in days while the y-axis scale is cumulative survival fraction. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival; JRH: John Radcliffe Hospital; UCLH: University College London Hospital.

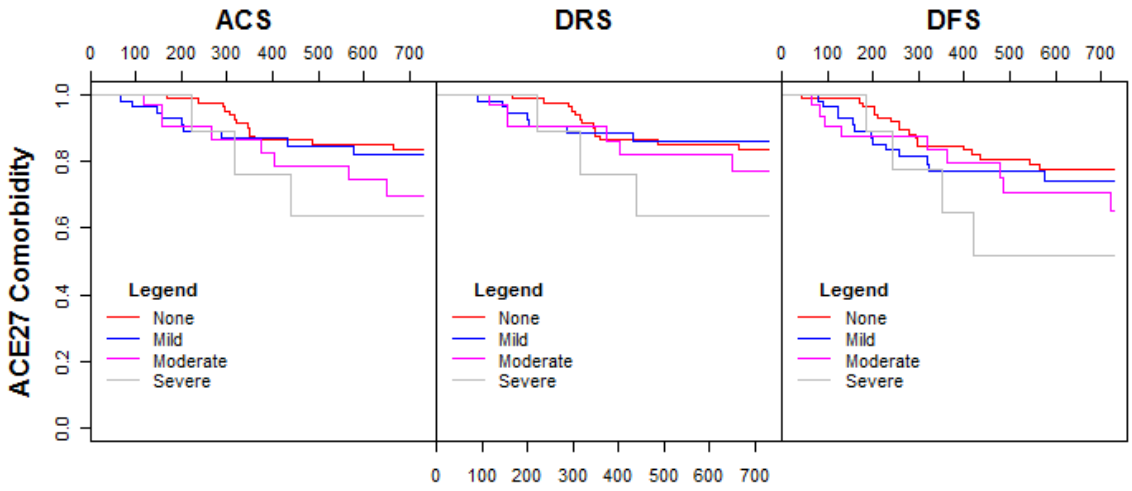


Figure 5.10: Survival by patient factors (2 of 2). Each square represents a Kaplan-Meier. The components involved each survival plot are indicated by the intersection of the patient factors (left margin of grid) and the type of survival (top margin of grid). The x-axis scale of each plot is in days while the y-axis scale is cumulative survival fraction. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival.

5.5 Tumour factors: clinical

This section presents the results and discussion for those factors determined by clinical inspection. Clinical factors include anatomic subsite and the growth pattern. Clinical staging is presented in Section 5.6.

Clinical tumour factors are summarized in Figure 5.11 and Figure 5.12, while results for OCLNM and survival are given in Figure 5.13 and Figure 5.14, found at the end of this section. Detailed information is available in Appendix IV.

5.5.1 Subsite and location

5.5.1.1 Results

Neither grouped subsite or location were significant predictors of OCLNM or survival. No conclusions can be drawn about individual subsites.

5.5.1.2 Discussion

The sample size was insufficient to perform any type of statistical analysis on individual subsites. While the number of cases of the oral tongue and floor were large, the sample sizes at other sites were insufficient. Based on previous studies it is expected that site-based differences in survival and metastasis do indeed exist [115, 121, 133]. The chosen grouping of subsites into upper oral cavity and lower oral cavity, though it has an anatomic basis, is ultimately arbitrary.

The oral tongue were given its own group because it was the most common subsite. The floor of the mouth is perhaps best grouped with the oral tongue, as the two are in continuity. However, with 31 cases (16%), the floor was second only to the oral tongue (97, 49%) for sample size. Combining the two would have left few cases for other groups.

The most relevant grouping would have been anterior versus posterior tumours. Most texts agree that risk of metastasis increases and prognosis worsens as tumour location becomes more posterior [3]. However, the database was constructed to identify subsite and not anterior-posterior position.

The argument could be made that grouping into upper and lower oral cavity makes sense based on the fact that the grouped tumours share common lymphatic drainage and regional metastasis patterns. However, the claim would be tenuous given the heterogeneity of the sites comprising each category. As such, results should be interpreted with caution.

Further, the designation of oral cavity subsite is subjective. This is especially true when tumours are large and span multiple subsites. In the patient record there was often disagreement about oral cavity subsite. This was also true of tumours that originated in the watershed areas between oral cavity subsites. When there was disagreement as to the subsite, the most common choice was used.

It was expected that tumours originating in the midline would be associated with decreased survival relative to their lateral counterparts, due to the associated risk of bilateral CLNM. This was not the case. However, the number of midline tumours was very small (n=11).

5.5.2 Growth pattern

5.5.2.1 Results

Growth pattern was not a predictor of OCLNM or survival.

5.5.2.2 Discussion

The results regarding tumour growth pattern were disappointing. Factors that can be easily evaluated pre-surgically are rare and valuable. In this study there was no evidence of correlation with survival or OCLNM. While this agrees with the results of Fukano et al. [81], it contradicts the findings of several other studies that note a significant relationship [3, 135].

The inadequacy of growth pattern likely extends from inconsistencies in the manner in which it was recorded. In many cases, the growth pattern was either not mentioned, or recorded in such a way as to make interpretation ambiguous. Exophytic tumours were sometimes described as "exophytic" by clinicians, but more often described by related terms like "raised". "Endophytic" was rarely used to describe tumours while "ulcerative" was used more frequently. Terminology like "invasive growth" or

"deep growth" were found more commonly, but represent a degree of subjectivity.

When the growth pattern was not conclusively described, the value was classified as "missing". Accordingly, there were a high number of missing growth patterns stemming from infrequent and non-standardized description. Compounding this problem was the existence of tumour subtypes that defy such categorical description. Superficial tumours can be plaque-like with no obvious exophytic or endophytic growth. Conversely, some tumours present as lesions with ulcerative central regions and exophytic peripheries.

Growth pattern alone is a poor choice for inclusion in a multivariate model, where sample size is paramount. The statistical results of growth pattern must be interpreted with caution. A formal evaluation of growth pattern at the time of clinical examination might be of value when determining pre-surgical prognosis, despite the findings of this study.

Figure 5.11: Distribution of oral cavity subsites in the patient sample

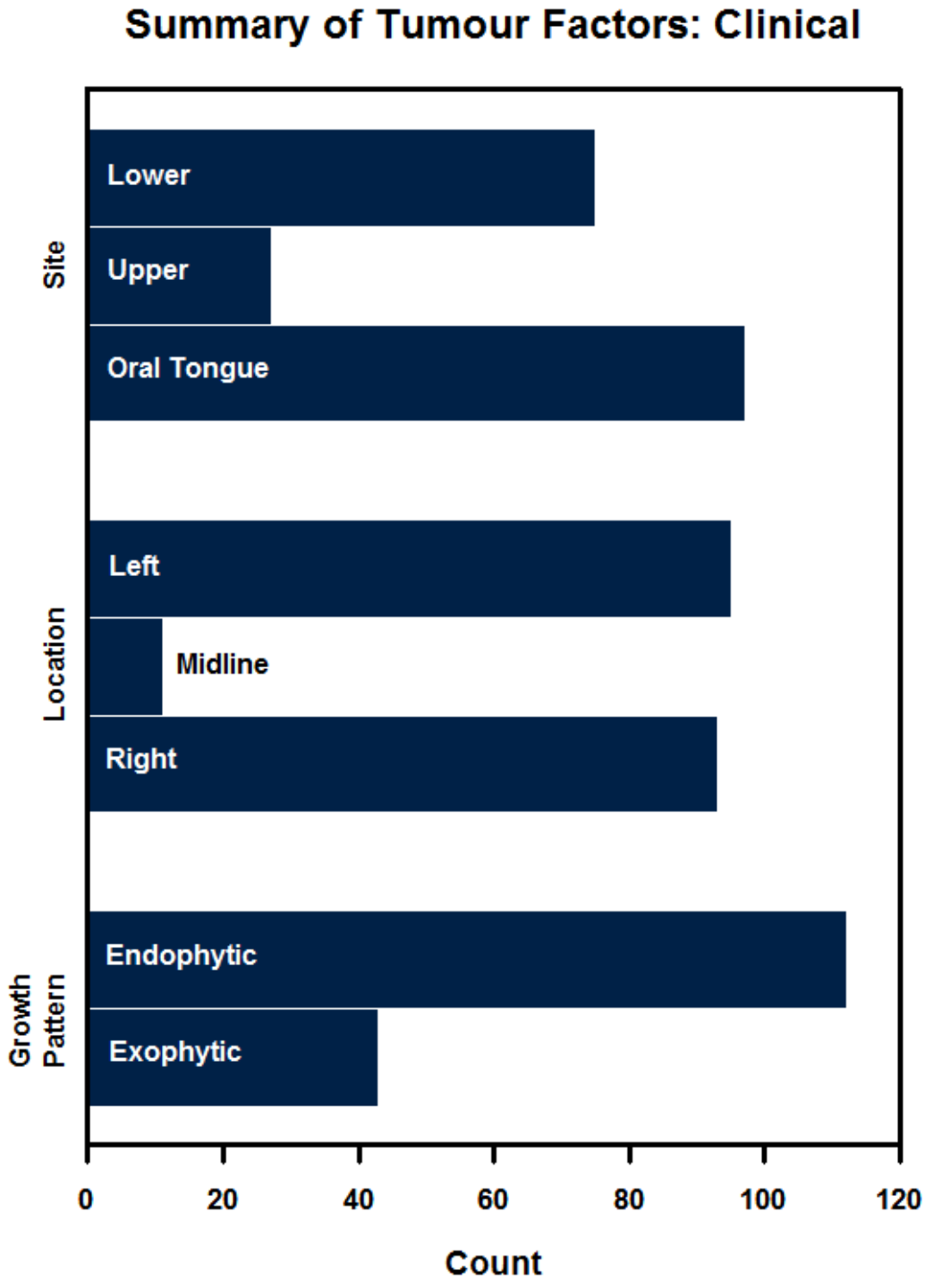


Figure 5.12: Distribution of clinical tumour factors.

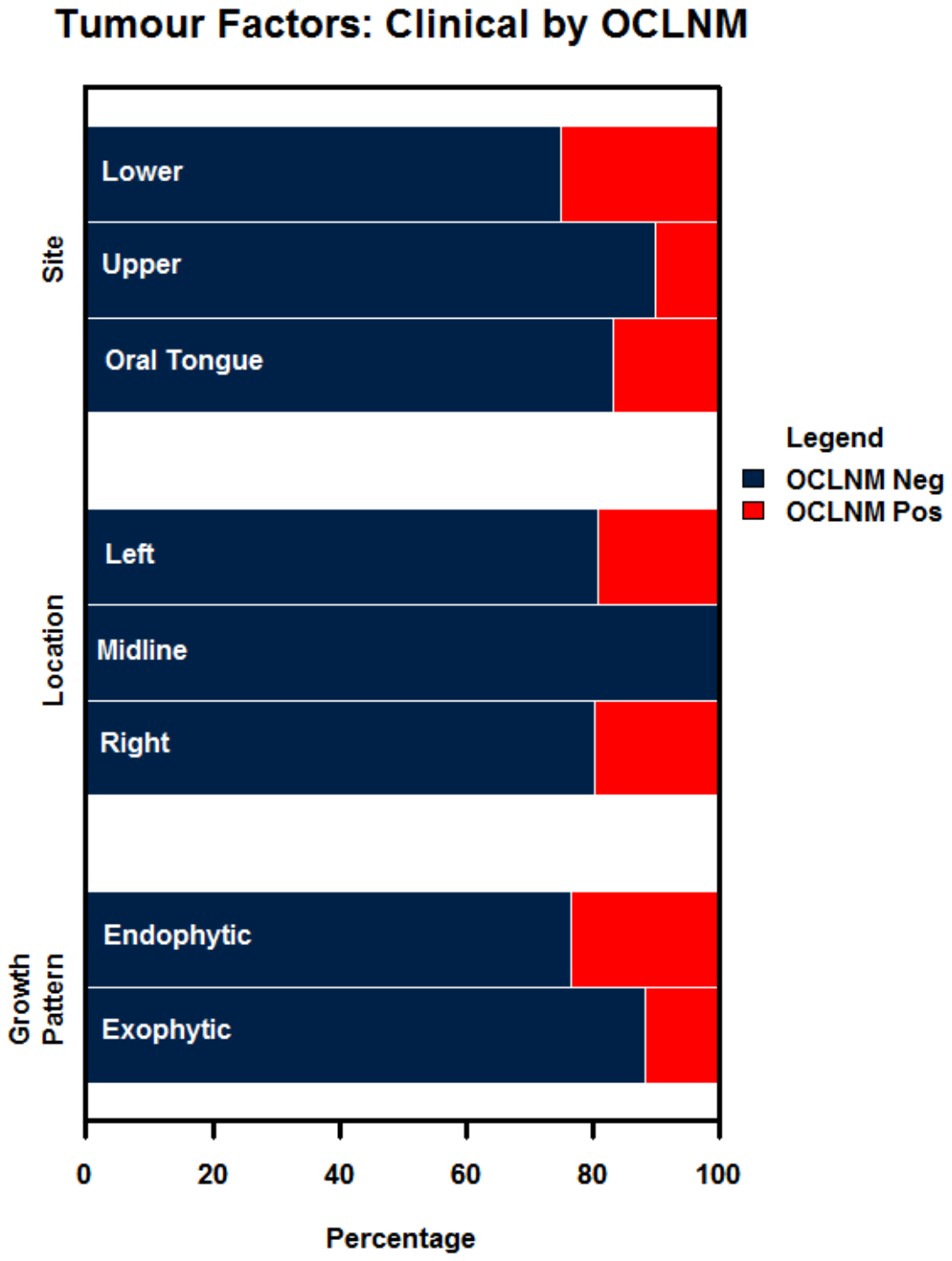


Figure 5.13: Clinical tumour factors by OCLNM status. Proportion of cases negative for OCLNM (blue) and positive for OCLNM are shown (red). OCLNM: occult cervical lymph node metastasis.

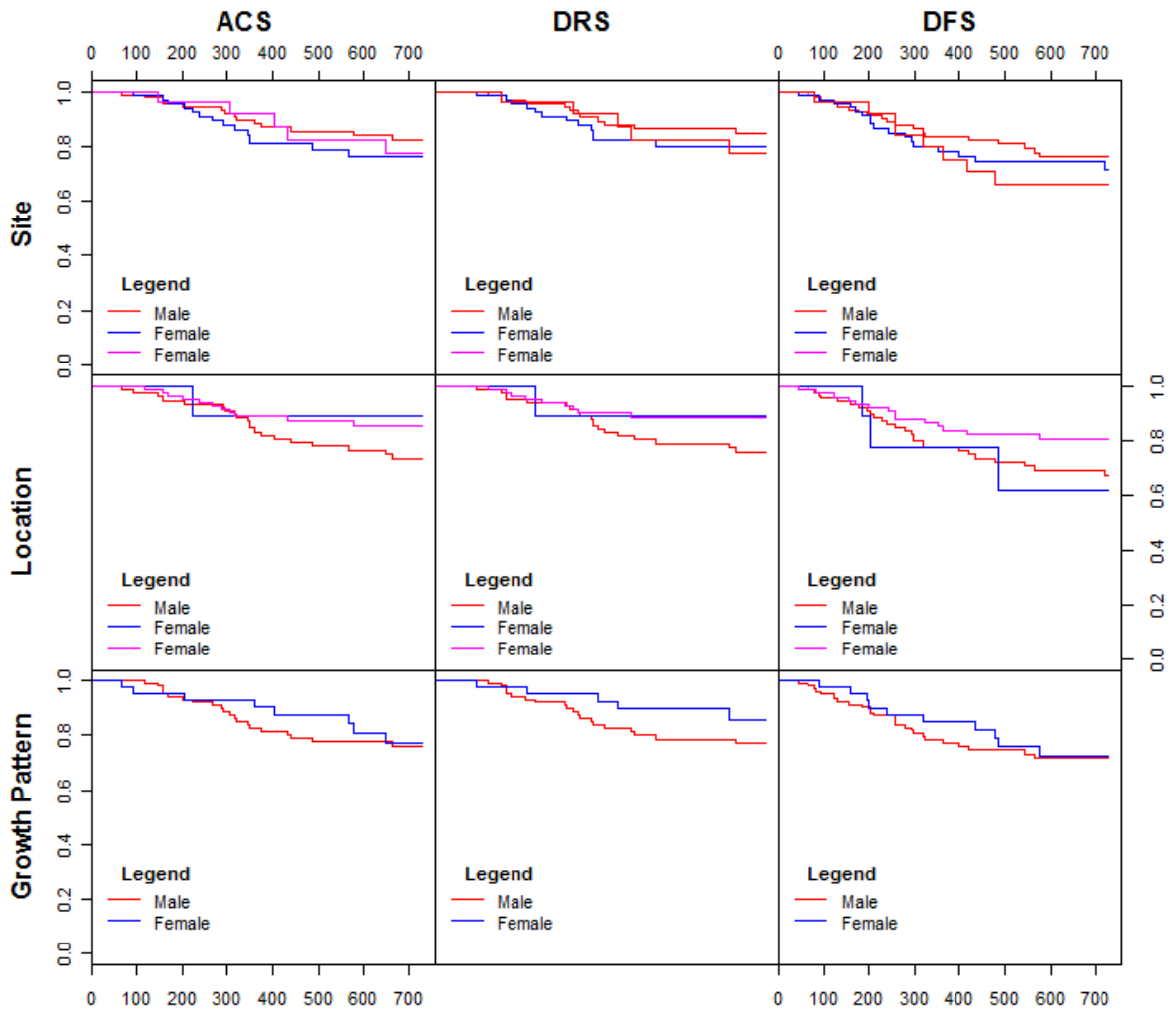


Figure 5.14: Survival by clinical tumour factors. Each square represents a Kaplan-Meier. The components involved each survival plot are indicated by the intersection of the clinical factors (left margin of grid) and the type of survival (top margin of grid). The x-axis scale of each plot is in days while the y-axis scale is cumulative survival fraction. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival; ORH: Oxford Regional Hospital; UCLH: University College London Hospital.

5.6 Tumour factors: staging

This section presents the results and discussion for those factors involved in disease staging. Factors include TNM criteria, CLNM distribution, extracapsular spread, bone invasion and midline invasion.

Staging tumour factors are summarized in Figure 5.15 - Figure 5.18, while results for OCLNM and survival are given in Figure 5.19 - Figure 5.23, found at the end of this section. Detailed information is available in Appendix IV.

5.6.1 T-stage

5.6.1.1 Results

cT4 and pT4 tumours had significantly elevated risk of OCLNM compared to cT1 and pT1 tumours respectively. pT4 tumours showed a trend towards significance.

cT4 tumours were significant predictors of ACS and DRS while there was a trend towards DFS. pT4 tumours were significant predictors of ACS, DRS and DFS. There was a trend towards significance with pT2 tumours and ACS.

5.6.1.2 cT vs. pT

When clinical and pathological classification were compared, the two agreed in only 118 (59.6%) cases. cT was up-staged by pT in 47 (23.7%) cases while cT was down-staged by pT in 33 (16.7%) cases. While there was a moderate correlation between cT and pT (Spearman = 0.66, $p<0.001$), the overall accuracy of cT was poor and cannot reliably predict pT. cT and pT are compared in Table 5.3.

Table 5.3: Comparison of cT and pT.

	pT1	*pT2*	*pT3*	*pT4*
cT1	*50*	*9*	*0*	*4*
cT2	*24*	*25*	*3*	*13*
cT3	*1*	*2*	*2*	*4*
cT4	*4*	*10*	*6*	*41*

5.6.1.3 Discussion

Only 9 (4.5%) cases were classified as cT3. This paucity of cT3 tumours has been noted by other authors [75]. cT3 tumours, defined as tumours exceeding a diameter of 40 mm, represent extensive local disease. Tumours this large likely invade critical structures, resulting in cT4 classification. The remaining 190 cases were relatively evenly distributed among cT1, cT2 and cT4 classifications (64, 65 and 61 cases respectively).

A similar distribution was seen for pT. Only 11 (5.6%) tumours were classified as pT3. The distribution of pT1 and pT2 tumours differed slightly, with 79 (39.9%) classified as pT1 and 46 (23.2%) pT2. The number of pT4 tumours was unchanged.

cT was a poor predictor of pT. The two agreed in only 60% of cases. cT is based on the clinical measurement of maximum surface diameter. In many cases it is difficult to distinguish malignant tissue from surrounding areas of dysplasia, exaggerating measurements. Measurement is also difficult when tumours occupy structures that are difficult to access, or have complex shapes like the buccal sulci. Pathology offers a more robust means for diameter measurement, though prognostic value is still questionable. It is not surprising that cT and pT tend to disagree.

cT stage was not a significant predictor of survival. There was a trend towards significantly reduced ACS and DRS when cT4 tumours were compared to cT1 tumours. However, in both ACS and DRS, survival was not significantly worse when cT2 or cT3 tumours were compared to cT1 tumours. Similar results were seen with OCLNM. Thus, clinical diameter is insufficient for classifying OCSCC.

The prognostic and predictive value of pT was slightly better than cT. ACS, DRS and DFS were significantly reduced when pT4 tumours were compared to pT1 tumours. Again, survival was not significantly reduced for pT2 and pT3 tumours. OCLNM behaved similarly, with only pT4 tumours having a significantly increased risk of regional metastasis.

The inability to T stage to predict survival or metastasis is not surprising and has been noted by

several authors [71-75]. This inadequacy is a major source of criticism for current TNM staging criteria. While the lack of prognostic significance for the T3 classification may be attributed to small sample size, the results still suggest that a diameter-based staging system should be reconsidered.

5.6.2 N-stage

5.6.2.1 Results

In 26 (13.1%) cases patients did not undergo neck dissection and were designated pNX. However, none of the 26 patients had a CLNM within a 6 month period after primary resection, or had post-surgical adjuvant therapy. They were, therefore, designated pN0.

cN1 and cN2 tumours were strong predictors of survival compared to cN0 tumours. Similarly, pN1 and pN2 tumours were strong predictors of survival.

5.6.2.2 cN vs. pN

The correlation between cN and pN was significant (Spearman=0.52, $p<0.001$); however cN correctly predicted pN in only 124 (62.6%) cases. 54 cases (27.2%) were up-staged, while only 19 (9.6%) were down-staged. cN and pN are compared in Table 5.4.

Table 5.4: Comparison of cN and pN.

	pN0	*pN1*	*pN2*
cN0	*110*	*8*	*16*
cN1	*19*	*11*	*30*
cN2	*0*	*0*	*3*

5.6.2.3 Discussion

The majority of CLNM positive patients were cN1 (60, 30.2%). Only 3 patients (1.5%) were classified as cN2 while none were classified as cN3. This is more likely a result of the inability of clinical examination and imaging to classify neck disease beyond "present" or "absent" rather than a lack of advanced metastatic disease.

On pathology, the number of patients classified as pN0 was relatively unchanged (129, 65.2%). There

was a major shift in the representation of pN1, with 20 (10.1%) cases, and pN2 disease, with 49 (24.7%) cases. Again, no pN3 cases reported. This highlights the enhanced ability of pathological analysis to distinguish between N-stage criteria compared to clinical assessment.

cN and pN agreed in 62% of cases. This is to be expected. Clinical examination and pre-surgical imaging of the neck have poor sensitivity and specificity for CLNM [136-138]. The poor sensitivity and specificity of MRI were also confirmed by the results. In this work the rate of OCLNM was 19%, lower than some studies that report values as high as 40% [146].

Even when clinical examination and imaging find evidence of CLNM, in the majority of cases they are unable to discriminate between cN1 and cN2 disease. As a result, a large number of tumours are upstaged by pathological analysis. In this work, upstaging occurred in almost 30% of cases.

Despite the shortcomings of cN classification, both cN1 and cN2 categories were strong predictors of ACS, DRS and DFS. With respect to ACS and DRS, approximately 30% fewer patients were alive at two years in the cN1 group compared to the cN0 group (RR 4.45 and 6.59 respectively). There were also 23% more cases of disease recurrence at two years in the cN1 group (RR 2.65). The results when cN2 was compared to cN0 are even more dramatic with relative risk of 25.8, 39.3 and 15.1 for ACS, DRS and DFS respectively. However, these results must be interpreted with caution as there were only three cases with cN2 disease.

The results of survival and occult metastasis analysis were equally dramatic for pathological analysis. Both pN1 and pN2 disease were highly significant predictors of survival and regional metastasis when compared to pN0 disease. The distinction between pN1 and pN2 disease is not pronounced. Regarding ACS, the relative risk of death was 7.73 for pN1 and 8.37 for pN2 disease when compared to pN0 disease. Mortality was 47% and 44% at two years for pN1 and pN2 disease respectively. Results were similar for DRS and DFS. This suggests that while CLNM itself is a strong predictor of survival and recurrence, little value is added by further N-classification.

5.6.3 Stage

5.6.3.1 Results

Clinically and pathologically stage III and IV tumours were significant predictors of survival compared to stage I tumours. Additionally, pathological stage II tumours showed a trend towards significance with ACS. For DRS, no statistics could be computed for pathological stage II tumours because no events were recorded.

5.6.3.2 cStage vs. pStage

Clinical stage predicted pathological stage in only 109 cases, or in 55.1% of cases. Clinical stage was increased by pathology in 50 cases (25.2%) and decreased in 38 cases (19.1%). Clinical examination and imaging correctly classified stage I tumours in 67.7% of cases, stage II in 59.3% of cases, stage III in 26.3% of cases and stage IV in 51.2% of cases. Despite the poor accuracy, there was a moderate correlation between clinical and pathological staging (Spearman 0.66, $p<0.001$). Clinical stage and pathological stage are compared in Table 5.5.

Table 5.5: Comparison of clinical stage and pathological stage.

	pStage I	*pStage II*	*pStage III*	*pStage IV*
cStage I	*44*	*4*	*2*	*6*
cStage II	*17*	*16*	*2*	*12*
cStage III	*2*	*2*	*5*	*24*
cStage IV	*2*	*5*	*10*	*44*

5.6.3.3 Discussion

The distribution of clinical stage was relatively uniform with the exception of Stage III with only 33 cases (16.6%). This is a product of the lack of cT3 tumours noted in an earlier section. There were even fewer pathological Stage II cases with 20 (10.1%). Additionally, there were relatively few Stage II tumours (27, 13.6%), leading to a concentration of tumours at the extremes, with 65 (32.8%) Stage I tumours and 86 (43.4%) Stage IV tumours.

Given the level of disagreement between cT and pT, and cN and pN, it is not surprising that the

clinical and pathological stage agreed only in about 55% of cases. While the predictive ability is considerably better from chance (chance being 25% for a four stage system) it remains far from the ideal value of 100% accuracy. While considerable progress has been made in clinical examination and imaging techniques, further substantial improvements must be made before there can be confidence in the ability to stage oral cancers pre-surgically.

Clinical staging had some value for predicting ACS, DRS and DFS. However, it is notable that there was no significant difference in survival or recurrence when stage II tumours were compared with stage I tumours. There was a decrease in survival and recurrence when stage III and stage IV tumours were compared to stage I tumours. Curiously, the relative risk of all-cause death, disease-related death and recurrence was higher in clinically stage III cancers versus clinically stage IV when both were compared to clinically stage I tumours (RR 7.32 vs. 4.27, RR 9.01 vs. 5.02 and RR 2.88 vs. 2.17 respectively). This trend must be interpreted with caution as the 95% confidence interval for each RR was large and it is unlikely that the difference was significant.

When histopathology was used to stage tumours, the results for ACS and DRS were similar to those of clinical staging. Stage II tumours were not significantly different with respect to neither ACS nor DRS. However, stage III and stage IV tumours were significant predictors of decreased survival when compared to stage I tumours for both ACS and DRS. Looking at the relative risks of each stage, a trend of increasing risk was apparent. For ACS, the relative risks of stage II, III and IV tumours were 2.50, 6.63 and 9.53. For DRS, although the relative risk for stage II tumours could not be calculated, the relative risks for stage III and IV tumours were 7.75 and 13.50 respectively. Regarding DFS, there was no significant difference in recurrence risk with stage II or stage III tumours, despite a significant difference with stage IV tumours. A trend of increasing relative risk was noted, though it was weaker compared to ACS and DRS (RR 0.89, 2.15 and 2.69 for stage II, III and VI respectively).

The results suggest that clinical and pathological staging of oral cancers does have some prognostic value. In this respect, pathological staging is superior and provides a good separation between

relative risk in each stage, though the relationship is less pronounced for DFS. The contribution of T-stage to the TNM scheme is questionable, and the overwhelming contribution of predictive power seems to come from N-classification. In this respect, changes to the T-stage criteria, if carefully performed, would be unlikely to have a detrimental effect on the prognostic value of TNM staging.

5.6.4 Cervical lymph node metastasis distribution

5.6.4.1 Results

MRI and pathology-detected CLNM were both strong predictors of survival.

5.6.4.2 Clinical CLNM vs. pathology CLNM

On MRI 22 of 58 (37.9%) pathology-confirmed cases of CLNM were detected. In pathologically N0 cases, MRI agreed in 62 of 123 (50.4%) cases. If we combine "equivocal" with "negative" cases to calculate a sensitivity and specificity for MRI-based CLNM detection, the values are 0.38 and 0.90 respectively. MRI and pathology are compared in Table 5.6.

Combined clinical examination and imaging identified metastasis in 44 of 58 cases (75.9%). The absence of metastasis was correctly identified in 111 of 129 cases (86.0%). In 24 of 134 cases (18.1%) identified as cN0, metastasis remained occult. Combined clinical and imaging evaluation had a sensitivity and specificity of 0.65 and 0.85 respectively.

Table 5.6: Comparison of detection of cervical lymph node metastases on MRI versus pathology.

		Pathology CLNM -	*Pathology CLNM+*
MRI CLNM -		*62*	*36*
MRI CLNM +		*61*	*22*
Sensitivity	*0.38*		
Specificity	*0.90*		

5.6.4.3 Discussion

25 patients (18.5%) had occult metastases. This is lower than literature reports of OCLNM ranging from 30-40% [44-47, 103]. The explanation is unknown, though improvements in detection techniques or reporter skill may be responsible. While imaging had a high number of false-negatives, false-positives were comparatively rare. When the combined clinical examination and imaging were

considered, sensitivity almost doubled to 0.65. It is clear that the combined efforts of imaging and clinical examination are superior to MRI alone.

Curiously, if one considers OCLNM as a separate entity from clinically detectable CLNM, there is a difference in survival between the two groups. In the present study, only 5 of 22 (22.7%) patients with OCLNM had all-cause mortality events, compared to 23 of 41 (56.1%) patients with clinically apparent metastases. The relative risk of all-cause mortality amongst those with OCLNM was 3.77 (relative to patients without metastasis), compared to 10.75 for those with clinically apparent metastasis. A similar relationship was seen with DRS and DFS analysis. It seems that while both groups had decreased survival, risk was less pronounced among patients initially classified as cN0.

The reason for this survival difference may be the fact that many occult metastases were too small to be found on clinical examination or imaging. Such metastatic deposits may represent early regional disease that is amenable to surgery and radiotherapy.

Overall, it seems that a combined clinical examination and MRI-based imaging approach provides valuable information on ACS, DRS and DFS. However, sensitivity remains poor and, as a result, the incidence of occult metastasis is high. While prognosis remains poor, the results suggest that patients with OCLNM have a survival advantage over those with clinically overt neck disease.

5.6.5 Extracapsular spread

5.6.5.1 Results

MRI ECS predicted DRS and DFS while there was a trend towards significance with ACS. Pathological ECS was a strong predictor of survival.

5.6.5.2 MRI vs. pathology

MRI found ECS in 5 cases. For the purposes of calculating sensitivity and specificity, the ability of MRI to detect ECS is considered in the 69 cases with pathology-confirmed CLNM. No cases negative for ECS were identified as positive by MRI for a specificity of 1.00. The sensitivity of MRI was 0.10 and

identified 46 of 51 positive cases as negative. MRI and pathology are compared in Table 5.7.

Table 5.7: Comparison of detection of the detection of extracapsular on MRI versus pathology.

	Pathology ECS -	*Pathology ECS +*
MRI ECS -	*18*	*46*
MRI ECS +	*0*	*5*
Sensitivity	*0.10*	
Specificity	*1.00*	

5.6.5.3 Discussion

The incidence of ECS in OCSCC patients undergoing primary surgical treatment was reported by Myers to be 38% [52]. This is less than the present case series in which 48 of 69 (69.6%) patients with pathology-confirmed CLNM had ECS. This may be due to different techniques used for the detection of ECS or differences in the patient populations served at the respective centers.

The ability of MRI to detect ECS was poor. Radiologists reported ECS in only 5 cases with a specificity of 1.00. However, the sensitivity was exceedingly poor at 0.10. As discussed in the previous section, MRI has poor sensitivity for detecting CLNM. It is not surprising that the sensitivity for ECS is much worse. In fact, 26 of 51 patients (51.0%) with pathology-confirmed ECS were not identified as having metastases by MRI. Second, MRI lacks the spatial and tissue resolution to identify small breaches of the node capsule. It may have been that much of the ECS reported on pathology was below the limits of MRI resolution. This also explains the lack of false-negatives. ECS must be substantial to be identified by MRI and few other conditions mimic its appearance.

Pathological ECS was a significant predictor of ACS, DRS and DFS. ACS and DRS were associated with 37% and 40% increases in 2-year mortality respectively. DFS was associated with a 32% increase in recurrence at 2-years. ECS detected by MRI was also a significant predictor of survival and recurrence, however this result must be interpreted with caution. Only 4 cases positive for MRI ECS were used for analysis.

5.6.6 Bone invasion

5.6.6.1 Results

Bone invasion detected by imaging was a significant predictor of OCLNM while bone invasion detected by pathology was not. Bone invasion on both imaging and pathology were predictors of ACS and DRS, but not DFS.

5.6.6.2 Imaging vs. pathology

Imaging detected pathological bone invasion with an accuracy of 91.3%. Imaging and pathology are compared in Table 5.8.

Table 5.8: Comparison of detection of bone invasion on MRI versus pathology.

	Pathology Bone -	*Pathology Bone +*
MRI Bone -	*138*	*13*
MRI Bone +	*4*	*39*
Sensitivity	*0.76*	
Specificity	*0.97*	

5.6.6.3 Discussion

Bone invasion was confirmed by pathological analysis in 45 of 152 (22.8%) available cases. It is difficult to know if this degree of bone invasion is normal or abnormal as published reports are rare.

There are few published results with which to compare the survival results. Soo et al. found a significant relationship between bone invasion and survival on univariate analysis that did not persist on multivariate analysis [149]. One might expect that the inclusion of bone invasion as a criterion for T4-stage by the AJCC implies that it is indeed of prognostic value. However, bone invasion was included in the original TNM criteria based on the consensus of the committee in charge of oral cavity staging [106]. Its inclusion has persisted despite little evidence to support or refute its value.

The results of this study partially support the consensus of the AJCC, at least when bone involvement is analyzed independently. However, the relationship is not strong, with a relative risk of approximately 2 for ACS and DRS. One might expect bone invasion to be a predictor of local

recurrence. The bone provides several low resistance paths for invasion, including the marrow cavity and mental canal of the mandible and the incisive foramen of the maxilla. However, this is not reflected in the results.

In the oral cavity subsites where bone invasion is common, such as the alveolar ridges, retromolar trigone and hard palate, little tissue intervenes between the bone and mucosa. Small tumours can invade bone simply by approximation and need not represent T4 disease. This undoubted weakens the relationship between bone invasion and survival and OCLNM

The ability of imaging to detect bone invasion was highly accurate at 91.3%. Specificity was remarkably high at 0.97. With these results, imaging-based assessment of bone invasion is a very useful surrogate for pathological assessment.

Given the high accuracy of imaging, it is not surprising that imaging-based bone invasion had similar results to pathological invasion with respect to survival. However, imaging-based bone assessment was a significant predictor of OCLNM. This may result from the tendency of imaging to report only cases of advanced invasion, while subtle bone invasion is likely to be reported as negative.

5.6.7 Midline involvement

5.6.7.1 Results

Tumours crossing or abutting the midline did not predict OCLNM. Tumours crossing the midline were predictors of ACS, DRS and DFS.

5.6.7.2 Discussion

The involvement of the midline can be evaluated by MRI imaging during pre-surgical staging procedures. Midline involvement is rare, with only about 13% of cases in the present study invading beyond its confines. This is likely because lateral tumours must be extensive before midline involvement occurs, and also because tumours originating in the midline are relatively rare.

Neither tumour abutting nor crossing the midline increased the risk of OCLNM. In theory, many of

the tumours involving or approaching the midline were extensive in local spread and, as such, an increased risk of OCLNM is expected. However, this was not seen. This is likely because of a lack of statistical power due to the relatively small number of tumours abutting (10 cases) or crossing (4 cases) the midline.

Tumour crossing the midline was a predictor of survival. This result was expected. However, tumour abutting the midline had no predictive value. This is difficult to interpret. All of the cases of tumours abutting the midline were in the oral tongue. Authors report that the lingual septum is not a barrier to spread, yet the existence of tumours that abut, but do not cross, the midline suggests that this is not the case [3]. There is a possibility that a less aggressive tumour subtype exists that is resisted by the fibrofatty lingual septum. However the "abutting" phenomenon might also be an artefact of peritumoural edema extending to, but not crossing, the lingual septum. Reduced statistical power due to the small sample may be to blame.

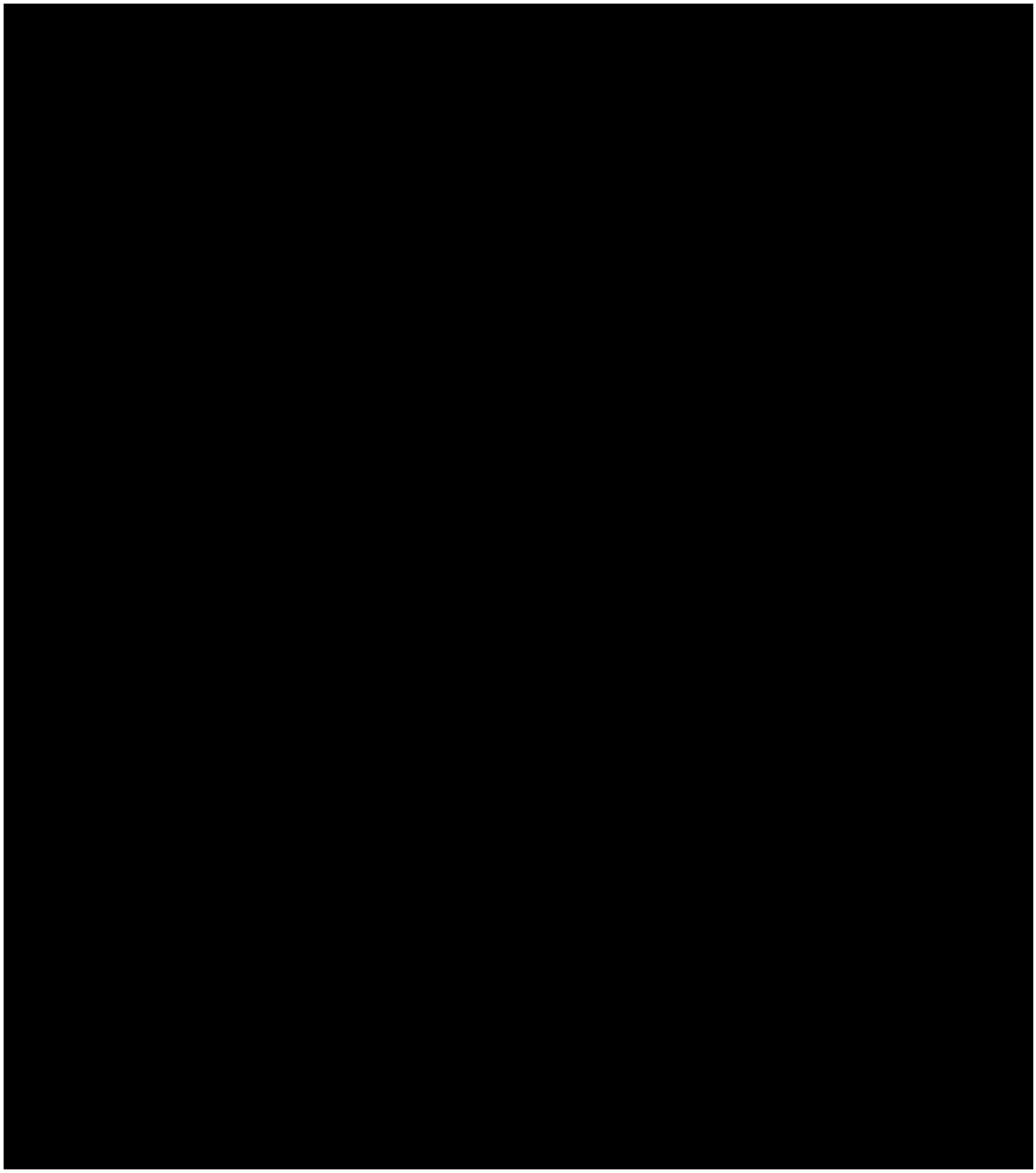


Figure 5.15 Distribution of cervical lymph node metastases by neck level for each oral cavity subsite (1 of 2). The coloring of each level represents the probability of metastasis at each level in the patient sample.

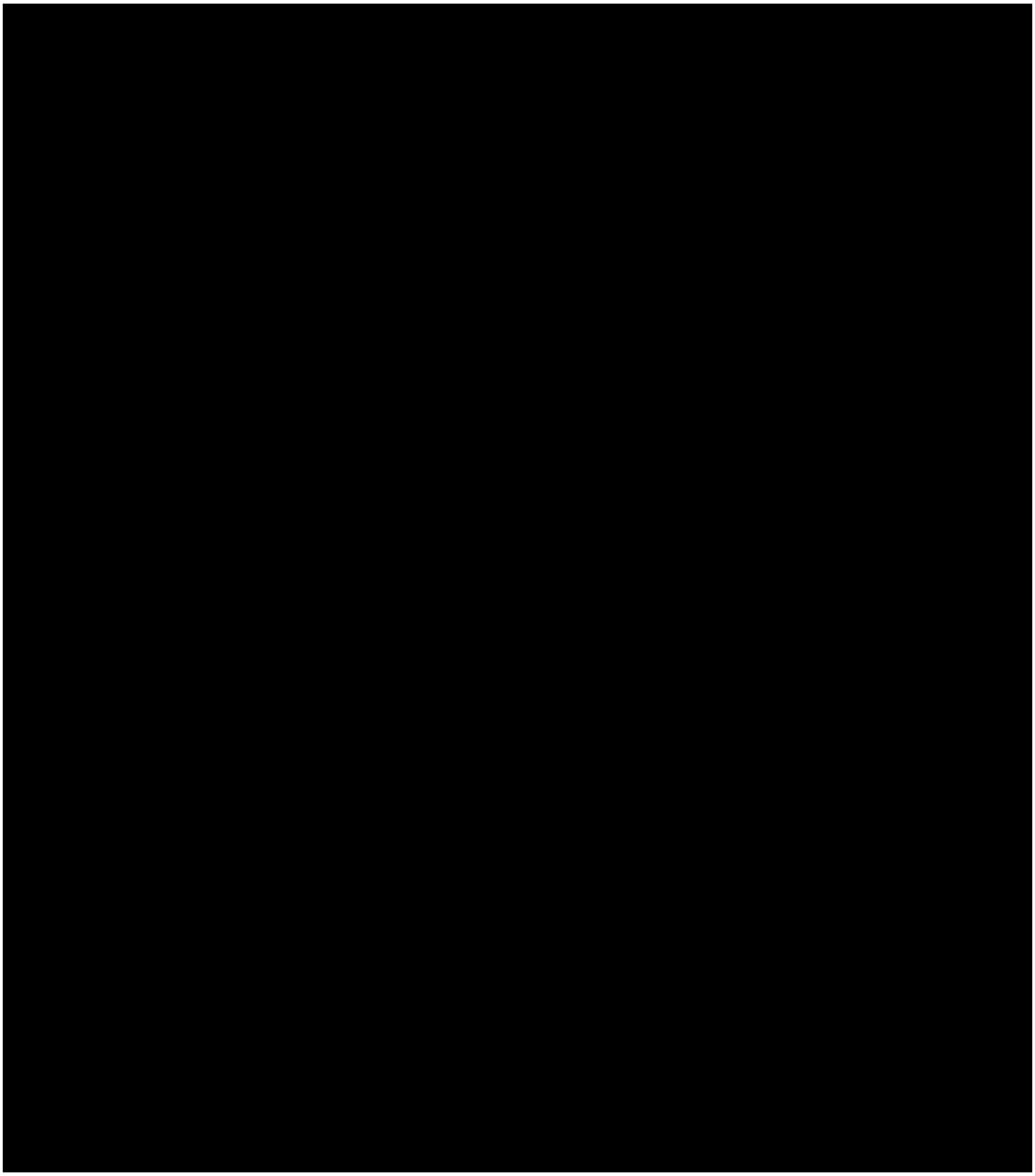


Figure 5.16 Distribution of cervical lymph node metastases by neck level for each oral cavity subsite (2 of 2). The coloring of each level represents the probability of metastasis at each level in the patient sample.

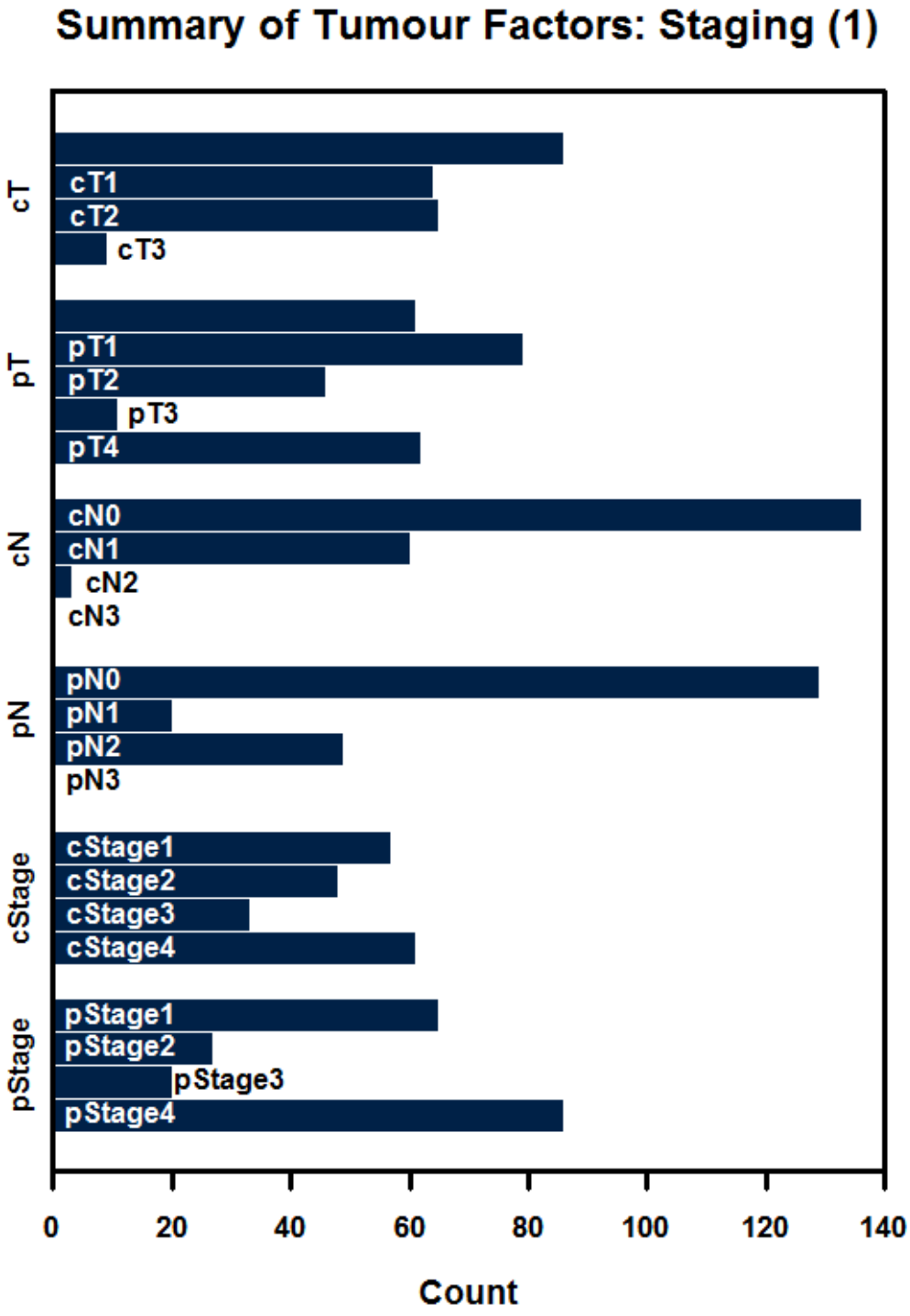


Figure 5.17: Distribution of staging tumour factors (1 of 2). The "c" prefix indicates clinical TNM staging while the "p" prefix indicates pathological staging.

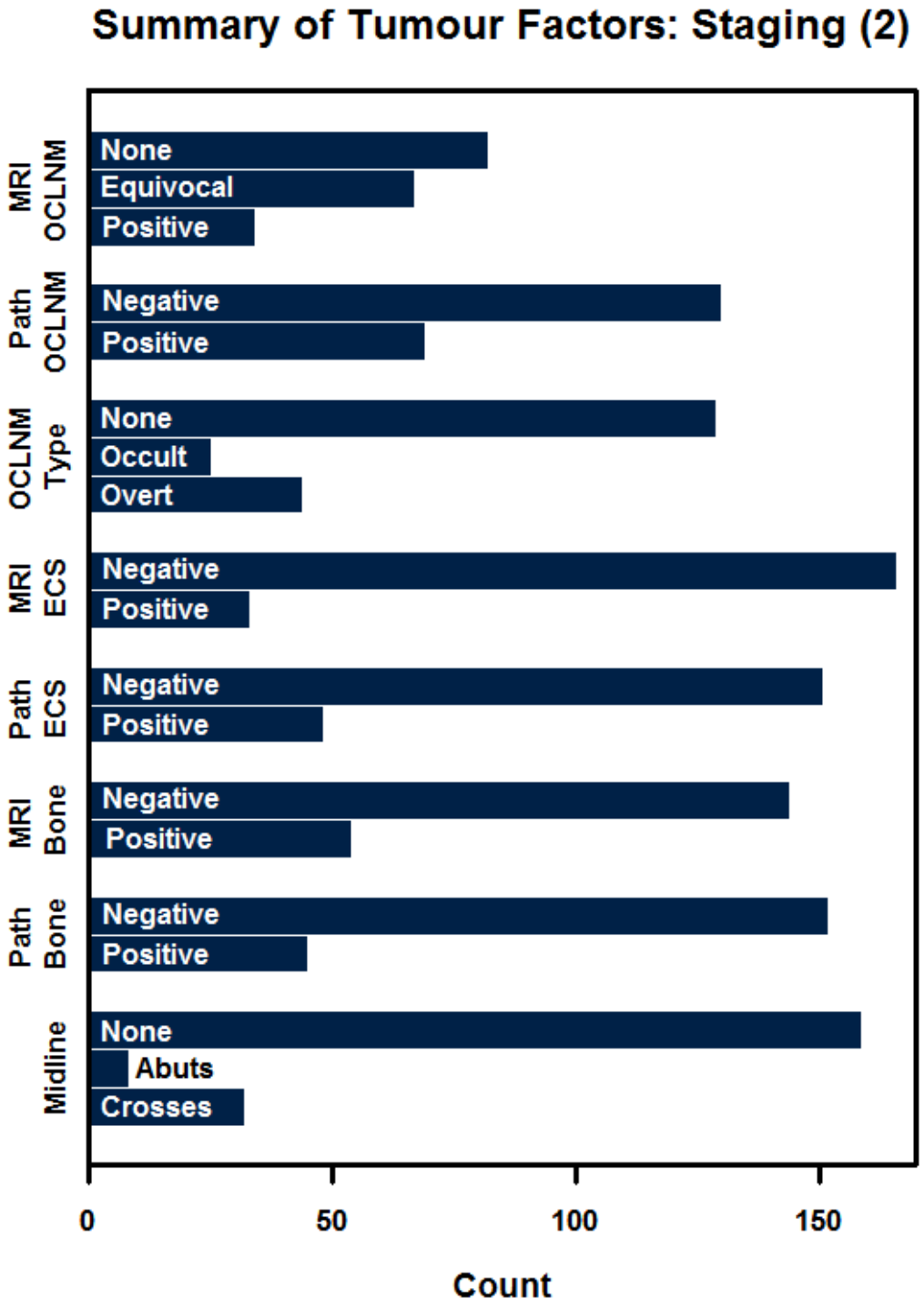


Figure 5.18: Distribution of staging tumour factors (2 of 2). OCLNM: occult cervical lymph node metastasis; Path: pathology; ECS: extracapsular spread; MRI: magnetic resonance imaging.

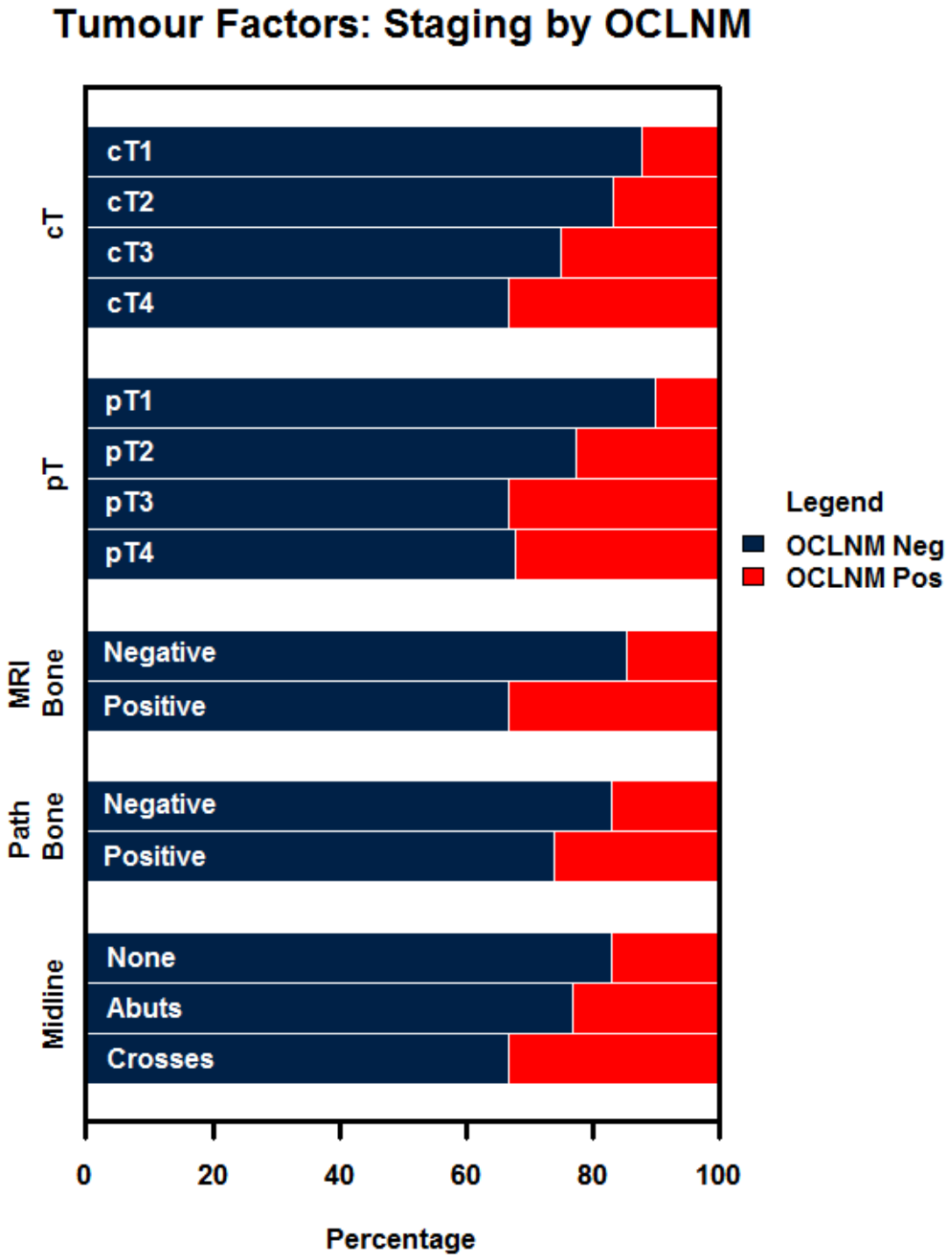


Figure 5.19: Staging tumour factors by OCLNM status. Proportion of cases negative for OCLNM (blue) and positive for OCLNM are shown (red). The "c" prefix indicates clinical TNM staging while the "p" prefix indicates pathological staging. OCLNM: occult cervical lymph node metastasis; Path: pathology; MRI: magnetic resonance imaging

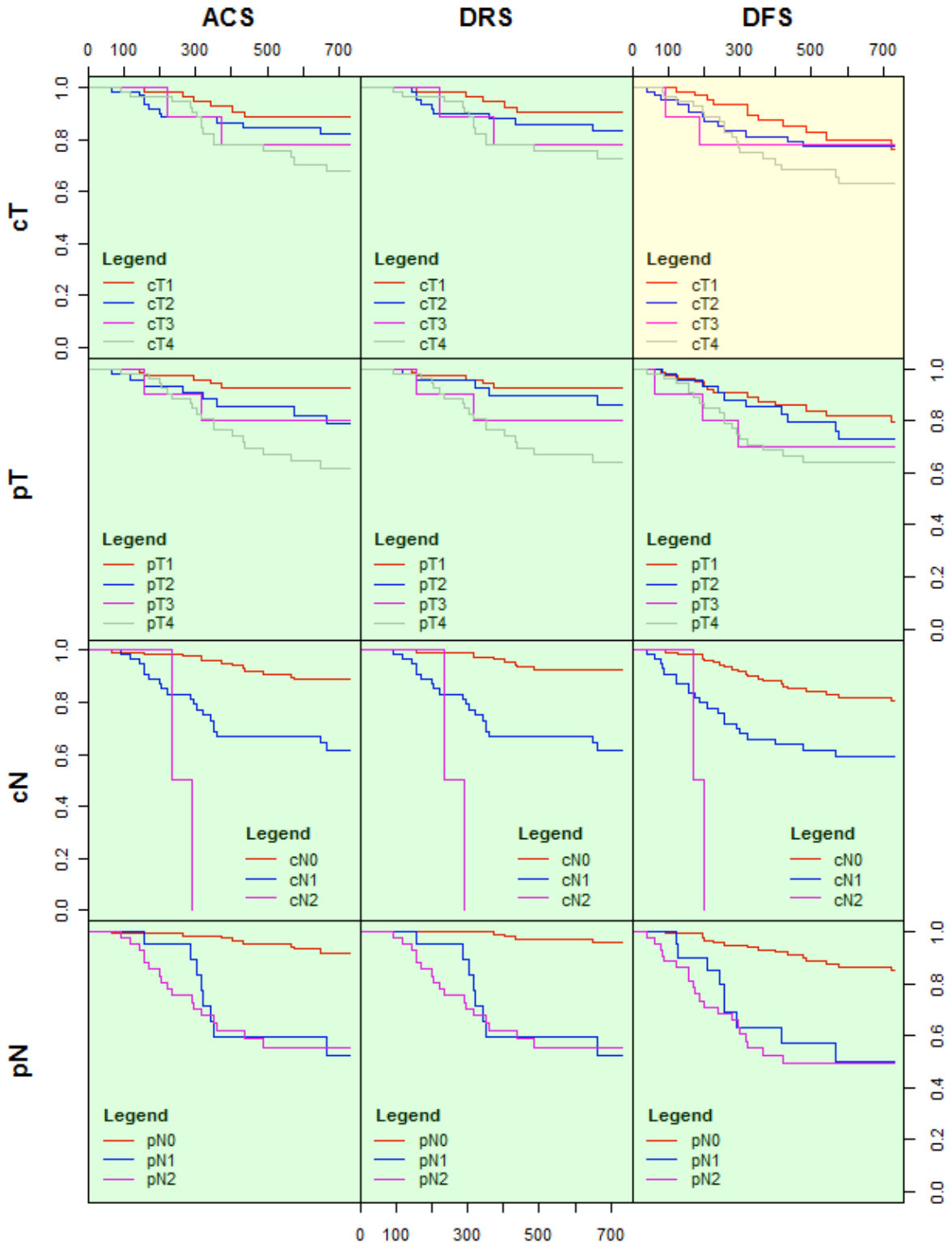


Figure 5.20: Survival by staging tumour factors (1 of 4). Each square represents a Kaplan-Meier curve. The components involved each survival plot are indicated by the intersection of the staging factors (left margin of grid) and the type of survival (top margin of grid). The x-axis scale of each plot is in days while the y-axis scale is cumulative survival fraction. The "c" prefix indicates clinical TNM staging while the "p" prefix indicates pathological staging. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival. Green indicates significant results. Yellow indicates trend towards significance (0.05 < p <= 0.10).

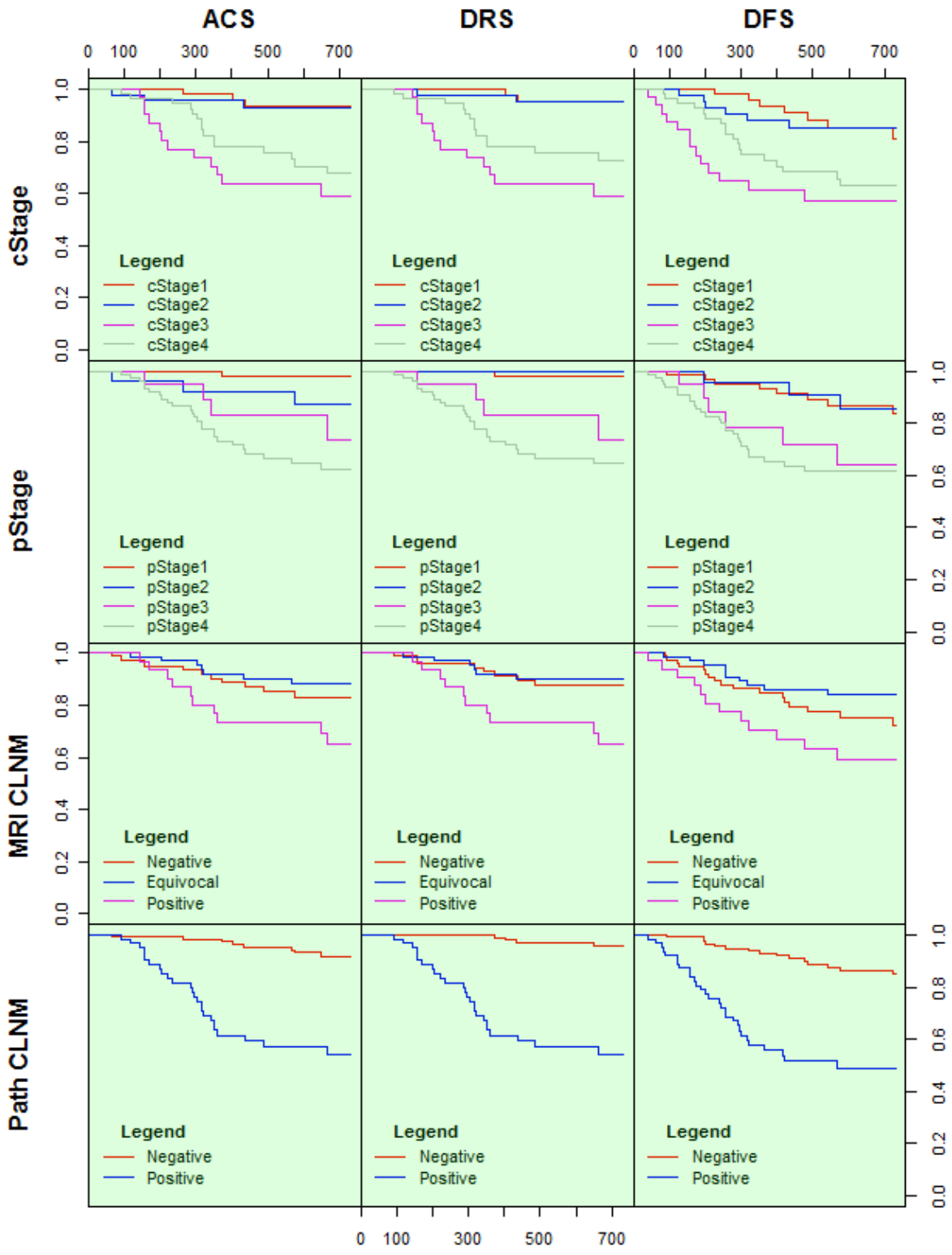


Figure 5.21: Survival by staging tumour factors (2 of 4). Each square represents a Kaplan-Meier curve. The components involved each survival plot are indicated by the intersection of the staging factors (left margin of grid) and the type of survival (top margin of grid). The x-axis scale of each plot is in days while the y-axis scale is cumulative survival fraction. The "c" prefix indicates clinical TNM staging while the "p" prefix indicates pathological staging. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival; path: pathology; MRI: magnetic resonance imaging; CLNM: cervical lymph node metastasis. Green indicates significant results.

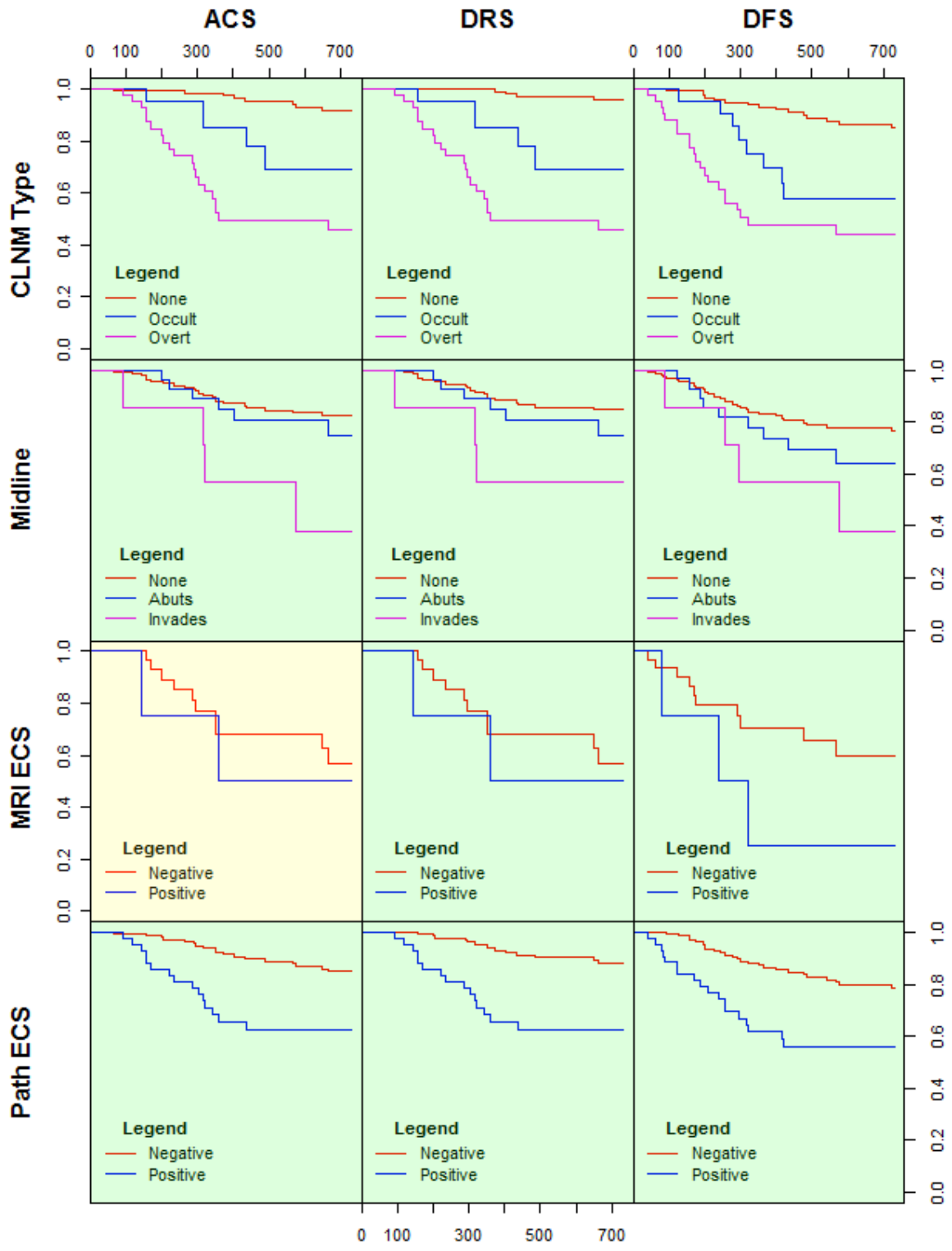


Figure 5.22: Survival by staging tumour factors (3 of 4). Each square represents a Kaplan-Meier curve. The components involved each survival plot are indicated by the intersection of the staging factors (left margin of grid) and the type of survival (top margin of grid). The x-axis scale of each plot is in days while the y-axis scale is cumulative survival fraction. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival; path: pathology; MRI: magnetic resonance imaging; CLNM: cervical lymph node metastasis; ECS: extracapsular spread. Green indicates significant results. Yellow indicates trend towards significance (0.05 < p <= 0.10).

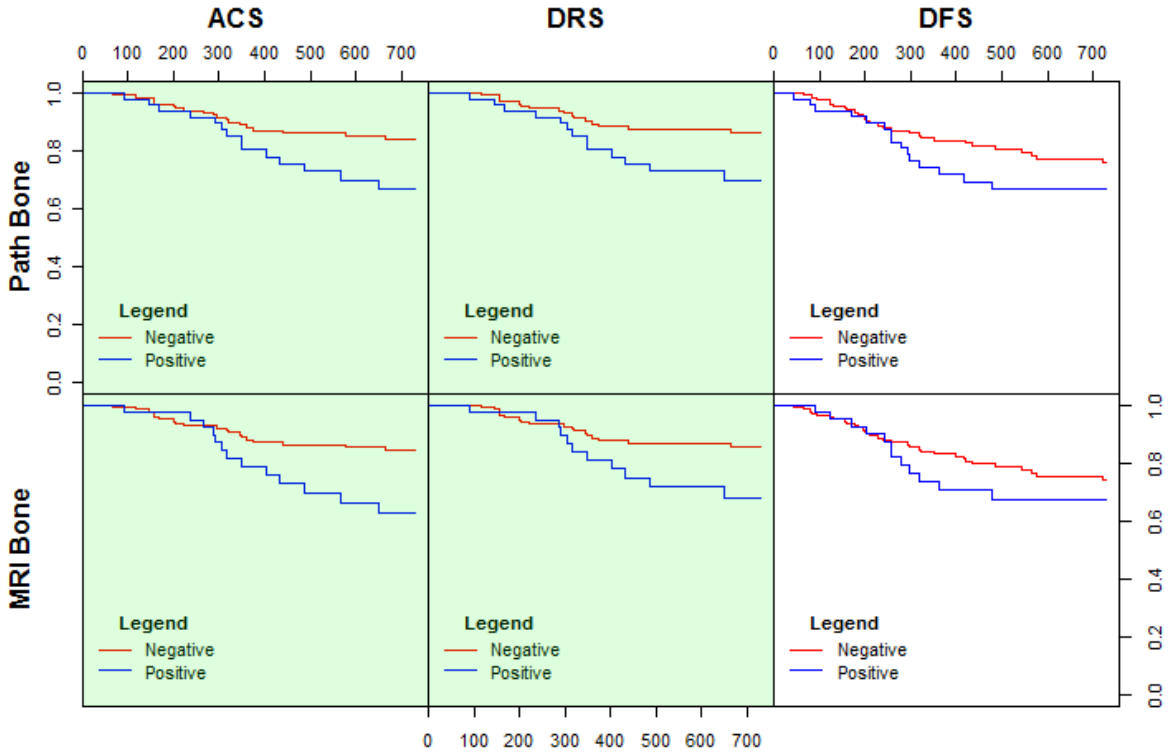


Figure 5.23: Survival by staging tumour factors (4 of 4). Each square represents a Kaplan-Meier curve. The components involved each survival plot are indicated by the intersection of the staging factors (left margin of grid) and the type of survival (top margin of grid). The x-axis scale of each plot is in days while the y-axis scale is cumulative survival fraction. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival; path: pathology; MRI: magnetic resonance imaging. Green indicates significant results.

5.7 Tumour factors: pathology

This section presents the results and discussion for those factors determined by pathological analysis Factors include tumour differentiation, infiltrative front, lymphovascular invasion, perineural invasion and surgical margins. TNM staging factors are discussed in Section 5.6.

Pathology tumour factors are summarized in Figure 5.24, while results for OCLNM and survival are given in Figure 5.25, Figure 5.26 and Figure 5.27, found at the end of this section. Detailed information is available in Appendix IV.

5.7.1 Tumour differentiation

5.7.1.1 Results

Poorly differentiated tumours were associated with increased risk of OCLNM and decreased survival.

5.7.1.2 Discussion

The majority of cases undergoing pathological analysis were moderately differentiated. Poorly and well differentiated tumours only account for 30.5% of cases. The subjective nature of the WHO grading scheme allows for considerable pathologist interpretation. The "moderate" grade acts as a "catch-all" for those tumours that are not clearly well or poorly differentiated. The Royal College of Pathologists state that this is to be expected but notes "it is important for prognostication to identify well differentiated and poorly differentiated tumours" [154].

The present study confirms that grade does have some importance. In the analysis, well differentiated tumours were used as the reference state. Moderate differentiation showed no significant difference in survival. However, it was apparent that patients with poorly differentiated tumours had decreased ACS, DRS and DFS. Similarly, moderately differentiated tumours had no increased risk of OCLNM; however risk was increased more than 8 fold among cases with poorly differentiated tumours.

This result is in agreement with several other authors [113-114, 155-157]. Poorly differentiated

tumours have little resemblance to the morphologic characteristics of normal epithelium and are known to be particularly aggressive and difficult to control [106].

5.7.2 Infiltrative front

5.7.2.1 Results

The infiltrative front did not predict OCLNM. However, non-cohesive tumour fronts were associated with decreased DRS and DFS with a trend towards significance with ACS.

5.7.2.2 Discussion

The non-cohesive invasive front was the most common pattern of invasion, present in 141 of 183 (77.0%) available cases. This suggests that the majority of OCSCC presents a particular challenge to surgeons trying to achieve local control as surgical margins cannot be reliably assessed visually.

The results of survival and OCLNM analysis are in agreement with several papers that suggest the infiltrative front is of prognostic and predictive importance [39-40, 160]. The non-cohesive front makes local control more difficult to achieve and may contribute to decreased survival. The presence of a non-cohesive front may also represent a more aggressive tumour phenotype with a predisposition for regional metastasis.

5.7.3 Lymphovascular and perineural invasion

5.7.3.1 Results

LVI was a strong predictor of OCLNM and survival. There was a trend towards significance for PNI and OCLNM. PNI was a significant predictor of ACS, trended towards significance with DRS and was not a predictor of DFS.

5.7.3.2 Discussion

Lymphovascular invasion was present in 41 of 197 (20.8%) available cases. This is consistent with published reports of the incidence of LVI [87, 117, 163].

It is likely that the reduction in survival, and increase in the rate of metastasis, stems more from the

association of the tumour with lymphatic vessels than blood vessels. Such an association provides the means by which cells can migrate to, and invade, cervical lymph nodes. The strong association between CLNM and survival may explain the increased mortality. It may also be that LVI is characteristic of a more aggressive tumour phenotype. The ability to invade vascular endothelium is acquired by accumulated mutation and may correlate with other harmful tumour characteristics. Blood vessel invasion is unlikely to directly contribute to decreased survival and increased CLNM. Hematogenous spread and associated distant metastasis are rare in oral cancer and seen only in advanced disease [144-145].

PNI was considerably more common than LVI, being present in 70 of 197 (35.5%) cases. This is consistent with other published reports [78, 172-174].

The lack of significant association of PNI with survival and OCLNM is surprising given the reputation of PNI as an "established" prognostic factor. Several authors note an association between survival [47, 78, 163, 170, 172-174, 176, 178, 180], recurrence [78, 84, 98, 163, 170, 173, 176, 178-179] and metastasis [98, 163, 170, 174-175].

The inability to predict local recurrence is particularly unexpected. In theory, invasion of the perineural space gives the tumour access to a low resistance route, enhancing local spread [163]. The Royal College emphasizes this aspect of PNI, saying that the presence of PNI "ahead of the invasive front" should be reported, while intratumoural PNI is "of doubtful prognostic significance" [154].

One possible explanation is the exclusion of SCC of the lip from the study sample. PNI is of particular prognostic importance in lip carcinoma. In the present study, the lack of prognostic and predictive value of PNI suggests that it is of less importance in the remainder of the oral cavity.

5.7.4 Surgical margins

5.7.4.1 Results

The condition of the surgical margin did not predict survival.

5.7.4.2 Discussion

The majority of patients in the study sample had either clear (45.9%) or close (42.9%) margins. That only 22 (11.2%) had involved margins is a testament to the skill of the surgeons involved.

The inability to predict survival is a disappointing result as there are several studies that show that involved surgical margins represent an increased risk of local recurrence and mortality [181-185]. This lack of association between the surgical margins and recurrence and survival is difficult to explain. It might be that the sample size is inadequate to detect any relationship. Only 22 patients had involved margins in the present study. It could also be that the 2-year survival period was insufficient and that differences would be detected with longer follow-up, although one would expect to see local recurrence early if disease persists. Finally, the only remaining explanation is that contemporary adjuvant therapies were sufficient to compensate for involved margins in the study sample, or at least reduce mortality and recurrence to levels comparable to those seen in cases with clear and close margins.

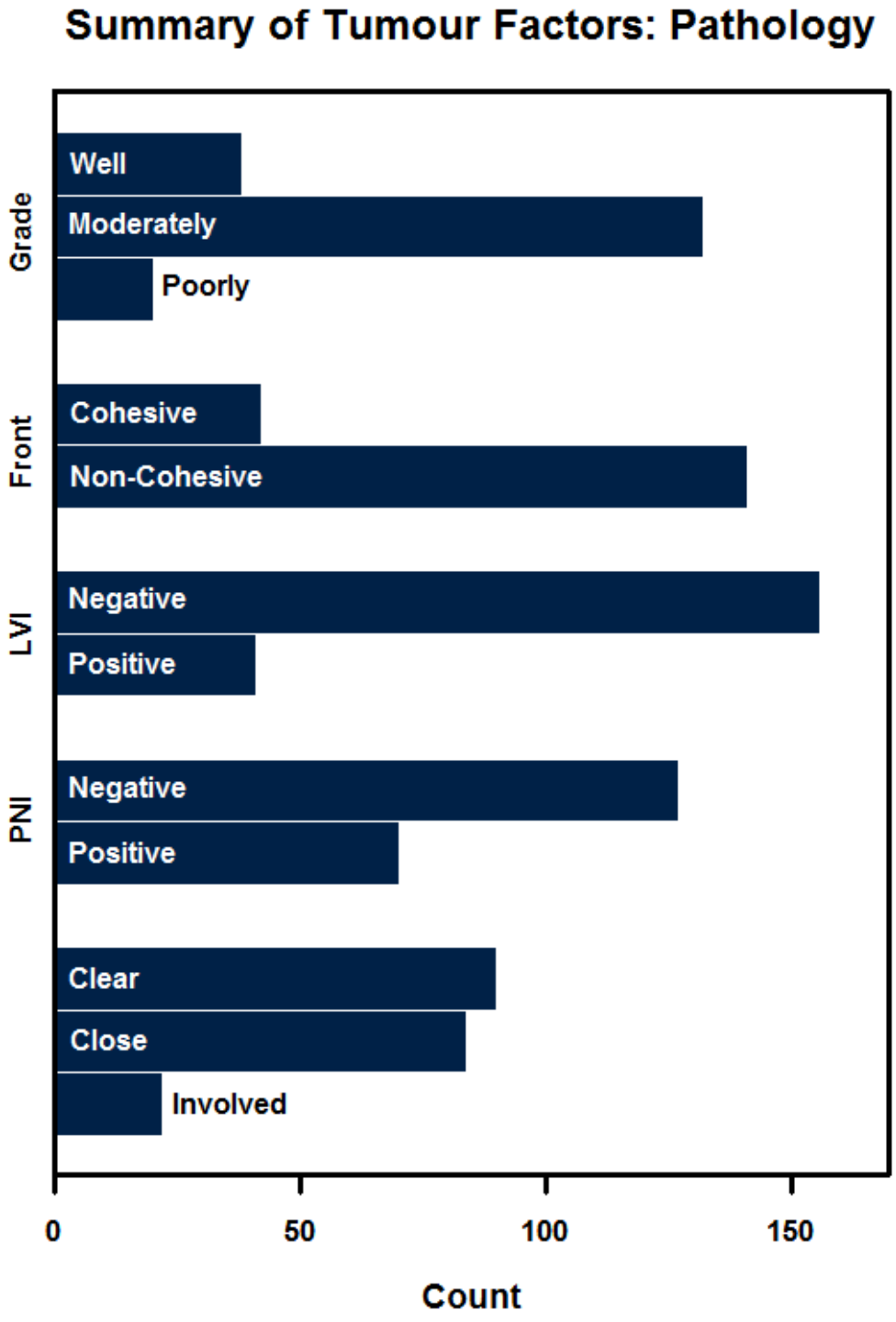


Figure 5.24: Distribution of pathological tumour factors. LVI: lymphovascular invasion; PNI: perineural invasion.

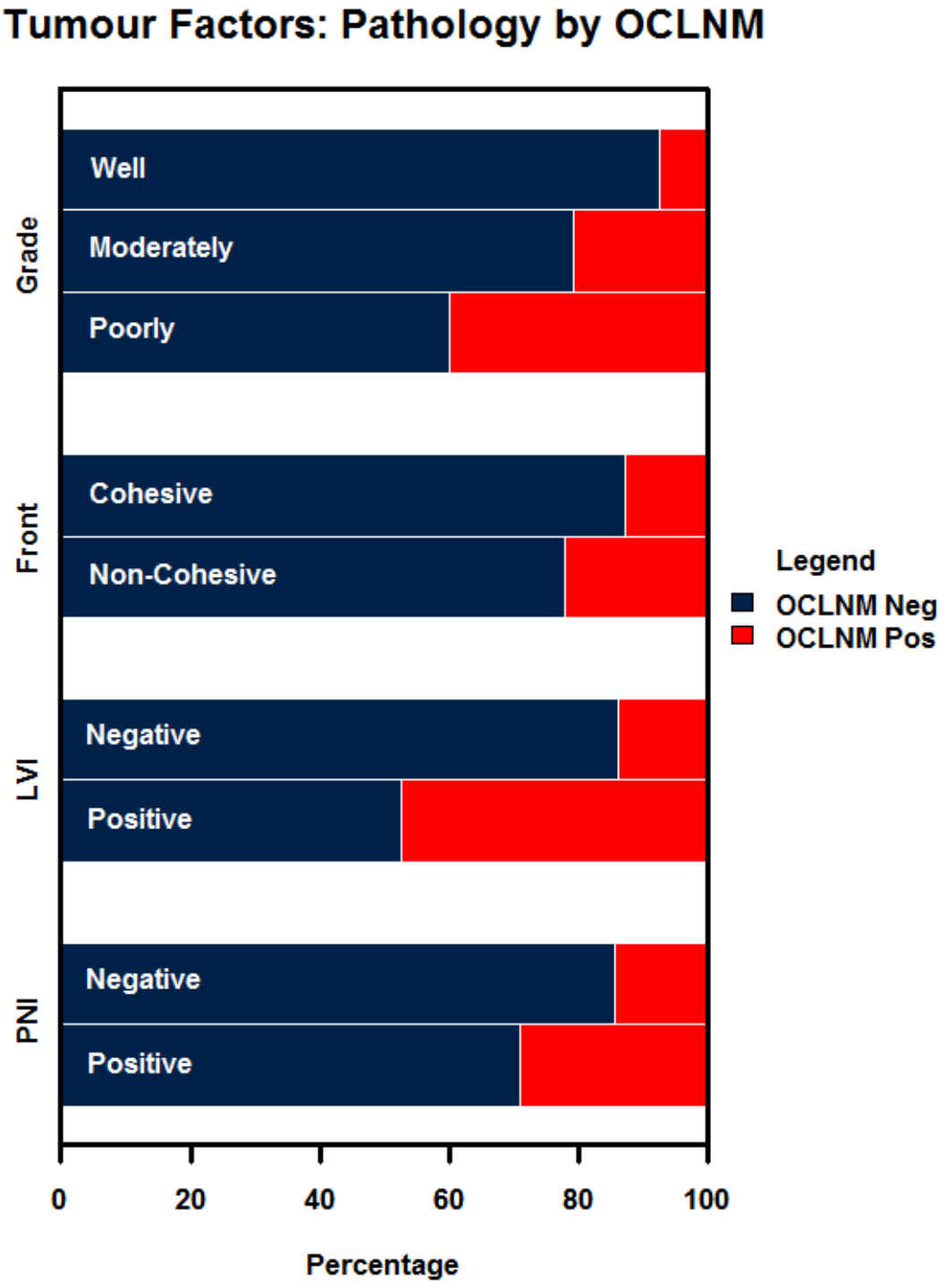


Figure 5.25: Pathological tumour factors by OCLNM status. Proportion of cases negative for OCLNM (blue) and positive for OCLNM are shown (red). OCLNM: occult cervical lymph node metastasis; LVI: lymphovascular invasion; PNI: perineural invasion.

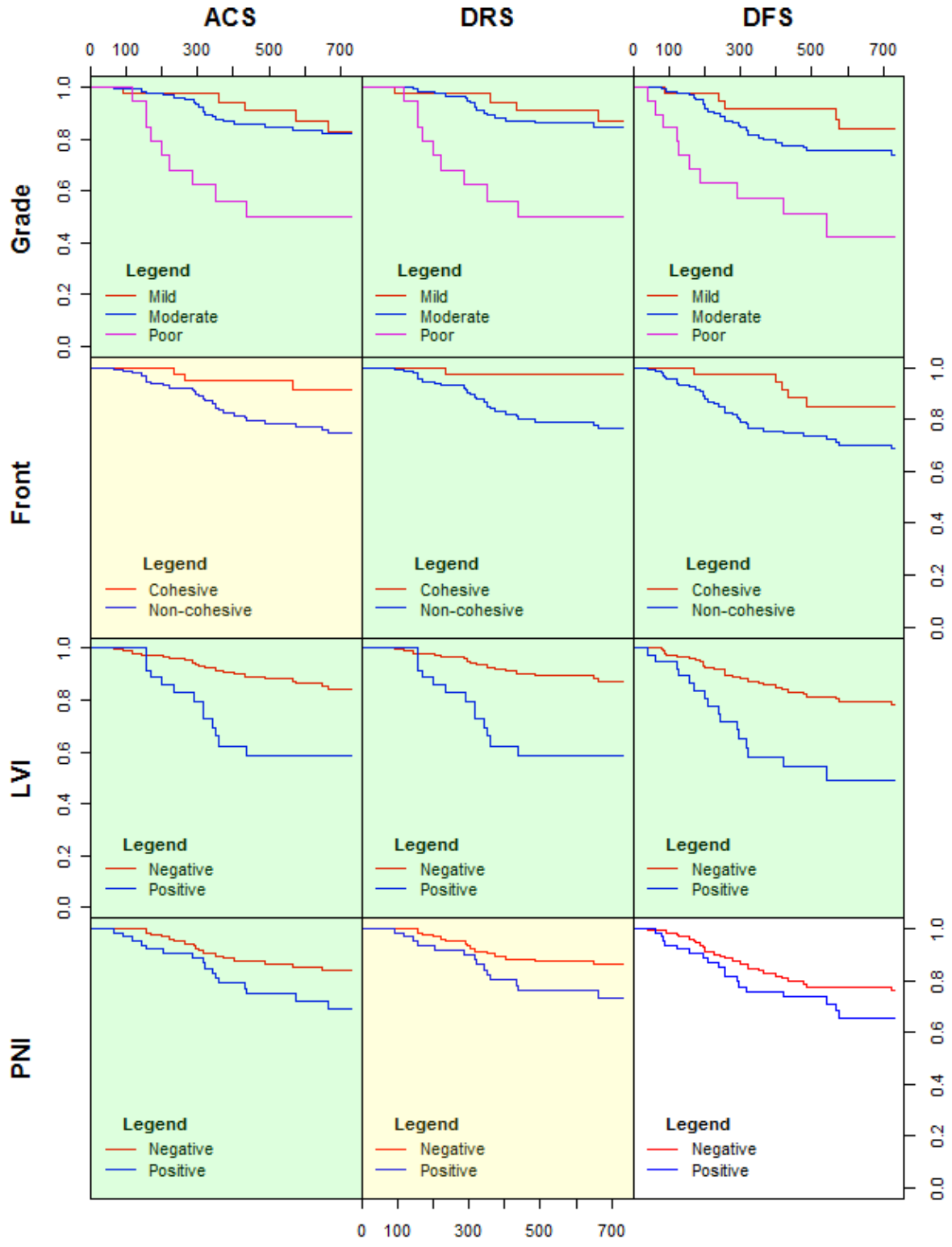


Figure 5.26: Survival by pathological tumour factors (1 of 2). Each square represents a Kaplan-Meier curve. The components involved each survival plot are indicated by the intersection of the pathological factors (left margin of grid) and the type of survival (top margin of grid). The x-axis scale of each plot is in days while the y-axis scale is cumulative survival fraction. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival; LVI: lymphovascular invasion; PNI: perineural invasion. Green indicates significant results. Yellow indicates trend towards significance (0.05 < p <= 0.10).

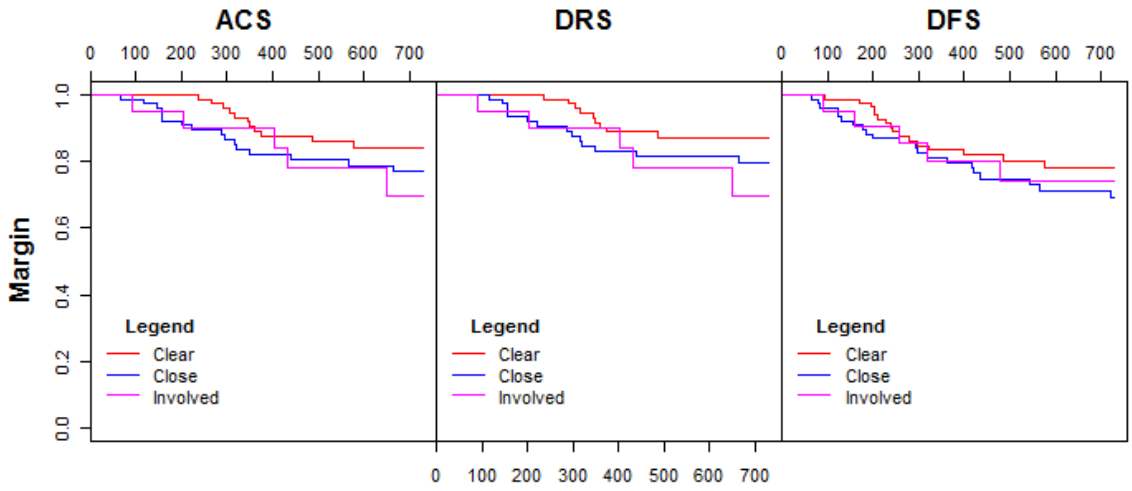


Figure 5.27: Survival by pathological tumour factors (2 of 2). Each square represents a Kaplan-Meier curve. The components involved each survival plot are indicated by the intersection of the pathological factors (left margin of grid) and the type of survival (top margin of grid). The x-axis scale of each plot is in days while the y-axis scale is cumulative survival fraction. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival.

5.8 Tumour factors: tumour dimensions

This section presents the results and discussion for those factors determined by MRI and pathology-based measurement. Factors include MRI-based volume, thickness and diameter, as well as pathology -based depth and diameter.

Staging tumour factors are summarized in Figure 5.28 - Figure 5.37, while results for OCLNM and survival are given in Figure 5.28 - Figure 5.42, found at the end of this section. Detailed information is available in Appendix IV.

5.8.1 Maximum diameter

5.8.1.1 Results

The intraobserver variability between the two diameter measurement sets was 0.950 (95% CI 0.933 - 0.962, $F = 39.5$, $p < 0.0001$), demonstrating that agreement between the two measurement sets was significant.

MRI diameter and log-transformed MRI diameter were strong predictors of OCLNM and survival. When categorized using the median value of 18.5 mm, MRI diameter was again a strong predictor of OCLNM and survival. When categorized using the first and third quartiles (0.0 mm, 28.5 mm), tumours exceeding 28.5 mm predicted OCLNM and DFS. ACS and DRS could not be assessed because there were no events in the less than 0.0 mm category.

Pathology diameter and log-transformed pathology diameter were strong predictors of OCLNM and survival. When categorized using the median value of 15.5 mm, pathology diameter was a predictor of OCLNM, ACS, DRS but not DFS. When categorized using the first and third quartiles (10.0 mm, 25.0 mm), tumours exceeding 25.0 mm predicted OCLNM, ACS and DRS but not DFS.

5.8.1.2 Discussion

MRI and pathology-based diameter were good predictors of OCLNM and survival. It was expected that diameter was of some value given its inclusion in the TNM staging criteria. However, recent

work suggests that depth of invasion is superior in these respects [71-73, 193].

Figure 5.28 and Figure 5.29 show that the distribution of MRI diameter is bimodal, due to the large number of T1 tumours designated as 0.0 mm. This issue is not solved by log transformation, with the same bimodal distribution persisting. Despite this, both the native and log-transformed data is highly predictive of OCLNM and survival when left in continuous form.

In this situation, it is justified to categorize MRI diameter, despite the ensuing loss of information. As discussed previously, doing so using the "lowest p" method leads to over-optimism, thus, in this work, the median was used for dichotomization and the first and third quartiles for division into three categories. Indeed, distinct survival groups are well defined by the MRI diameter categories. The three category method, in particular, provide good separation (seen in Figure 5.40) for groups in ACS, DRS and DFS. While ACS and DRS could not be assessed statistically, visual inspection reveals that MRI diameter is useful in this respect.

The results of pathology-base diameter reveal much the same. However, when grouped based on the median and the first and third quartiles, pathology-based diameter did not predict DFS. As well, curve separation for the pathology diameter groupings is less pronounced.

It is unclear why pathology-based diameter appears to be inferior to MRI-based diameter. MRI overestimates tumour dimensions due to peritumoural edema, while pathology underestimates tumour dimensions due to shrinkage during formalin fixation [67, 100-101]. It could be that MRI-based measurement has a component that is based on the severity of tumour-associated inflammatory reaction that adds to its predictive value.

MRI-based diameter measurement carried a degree of uncertainty due to the methods used. The diameter was measured in either the coronal or axial plane. There is a possibility that the maximum diameter was outside of those planes and thus not seen. As well, the diameter in one of the two planes evaluated was based on the number of tumour-containing slices. Thus, in these cases there

was an error factor associated with diameter proportional to the slice thickness and spacing.

Pathology diameter measurement was carried out by trained technicians in the respective pathology departments at UCLH and the JRH. While these institutions follow national standardized protocols, it may have been that measurement technique varied. This is impossible to assess.

5.8.2 Depth of invasion and thickness

5.8.2.1 Results

The intraobserver variability between the two thickness measurement sets was 0.984 (95% CI 0.979 - 0.988, F = 127.0, p < 0.0001), demonstrating that agreement between the two measurement sets was significant.

MRI thickness and log-transformed MRI thickness were strong predictors of OCLNM and survival. When categorized using the median value of 11.0 mm, MRI thickness was again a strong predictor of OCLNM, ACS and DRS with a trend towards significance in DFS. When categorized using the first and third quartiles (4.0 mm, 16.0 mm), tumours exceeding 16.0 mm predicted OCLNM, ACS and DFS. There was a trend towards significance in OCLNM, ACS and DFS for tumours in the intermediate category. DRS could not be assessed because there were no events in the less than, or equal to, 10.0 mm category.

Pathology depth and log-transformed pathology depth were strong predictors of OCLNM and survival. When categorized using the median value of 6.0 mm, pathology depth was a predictor of ACS, DRS and DFS but not OCLNM. When categorized using the first and third quartiles (3.0 mm, 10.0 mm), tumours exceeding 10.0 mm predicted OCLNM and DFS. ACS and DRS could not be assessed because there were no events in the less than or equal to 3.0 mm category.

5.8.2.2 Discussion

MRI-based thickness and pathology-based depth were strong predictors of OCLNM and survival. This is in line with recently published data that depth is an important predictor of OCLNM and survival

[193]. However, most of the published work is with pathological samples. Work with MRI-based measurement is much rarer, and results are mixed [102-104]. The present work is one of the largest available with positive results on MRI-based thickness measurement.

Like MRI-based diameter, the bimodal distribution of thickness made analysis difficult. However, both native and log-transformed continuous data were strong predictive and prognostic factors. Nevertheless, the same argument arises in favor of categorizing MRI thickness, despite the ensuing loss of information. Again, distinct survival groups are well defined by the MRI thickness categories. The three category method provided good separation (seen in Figure 5.40) for groups in ACS, DRS and DFS. While DRS could not be assessed statistically, visual inspection reveals that MRI thickness is useful in this respect.

Pathology-based diameter had similar results to MRI-based thickness when left as a continuous variable. When dichotomized at the median, depth was unable to predict OCLNM, but was a predictor of survival. When divided using the first and third quartiles tumours exceeding 10.0 mm were at higher risk for OCLNM and had reduced DFS. Though ACS and DRS could not be assessed, visual inspection of the Kaplan-Meier curves show that thickness provides good separation. However, in all three survival categories, the separation between categories is not as pronounced as it is with MRI-based thickness.

Pathology-based tumour depth of invasion is quickly becoming the favored measurement for predicting OCLNM and survival. However, these results suggest that MRI-based tumour thickness may be superior to pathology-based depth. As with diameter, there are problems associated with both imaging and pathology methods and it is unclear why MRI should be superior. It must be pointed out that the difference between MRI and pathology was small and, given the small sample size, they provide similar results.

By definition thickness differs from depth of invasion. Depth is based on a reconstructed mucosal surface, ignoring exophytic growth and ulcerative defects. Tumour thickness measures from the

surface of the malignant tissue and thus overestimates depth in exophytic growths and underestimates depth in ulcerative tumours. However, considering that the tumour will conform to the contours of the oral cavity structures when the mouth is closed, attempting to reconstruct the mucosal barrier is a fruitless exercise on MRI.

5.8.3 Volume

5.8.3.1 Results

The intraobserver variability between the two volume measurement sets was 0.987 (95% CI 0.982 - 0.990, $F = 149.0$, $p < 0.0001$), demonstrating that agreement was significant.

MRI volume displayed a trend towards significance for OCLNM prediction and was a strong predictor of survival. Log-transformed MRI volume was a strong predictor of OCLNM and survival. When categorized using the median value of 3.0 cm^3, MRI volume was again strong predictor of OCLNM and survival. When categorized using the first and third quartiles (0.5 cm^3, 8.0 cm^3), tumours exceeding 8.0 cm^3 predicted OCLNM and survival. Tumours in the intermediate category also predicted OCLNM, ACS, DRS but not DFS.

5.8.3.2 Discussion

The results of MRI volume demonstrate the potential of this metric as a candidate for refined T-staging criteria. Volume was a strongly significant predictor of OCLNM and survival as a continuous variable.

Like MRI-based diameter and thickness, volume was rightward skewed and bimodal. The rightward skew was more pronounced with volume, owing to the fact that it is a three-dimensional measurement, compared to the one-dimensional diameter and thickness. As such, it is likely that categorization is a viable option. When dichotomized, volume was again a strong predictor of OCLNM and survival. When divided into three categories, both tumours exceeding 8.0 cm^3 and tumours in the intermediate category were significant predictors of OCLNM, ACS and DRS. This is different from all other categorized measurement variables in which only the extreme grouping displayed any

difference in OCLNM or survival. This points to the fact that MRI-based volume may be superior to MRI diameter, MRI thickness, pathology diameter and pathology depth for the evaluation of CLNM and prognosis. However, a similar caveat also applies. Tumour volume was only slightly better than other predictors and, thus, it is difficult to conclude anything beyond that the predictors were similar. It is possible that multivariate modeling will reveal more about the value of each measurement variable.

The literature on MRI-based volume measurement is small. Yuen found no relationship between MRI-measured volume and survival [100]. Similarly, Preda found no such relationships [101]. Chew et al., however, had more success [105]. The group found that volume, using a cut-off value of 13.0 cm^3 was predictive of ACS and DFS. The sample size used was only 17 cases of lingual SCC.

MRI-based volume measurement likely suffers from systematic error beyond that previously discussed for diameter and thickness. The relatively large slice thickness and spacing of MRI, and the resulting non-isometric voxels, mean that the volume derived is only an approximation. Accuracy would improve with decreased slice thickness, however MRI imaging has a long way to go before it reaches the isometric voxels of CT. Therefore, volumes must be interpreted with caution.

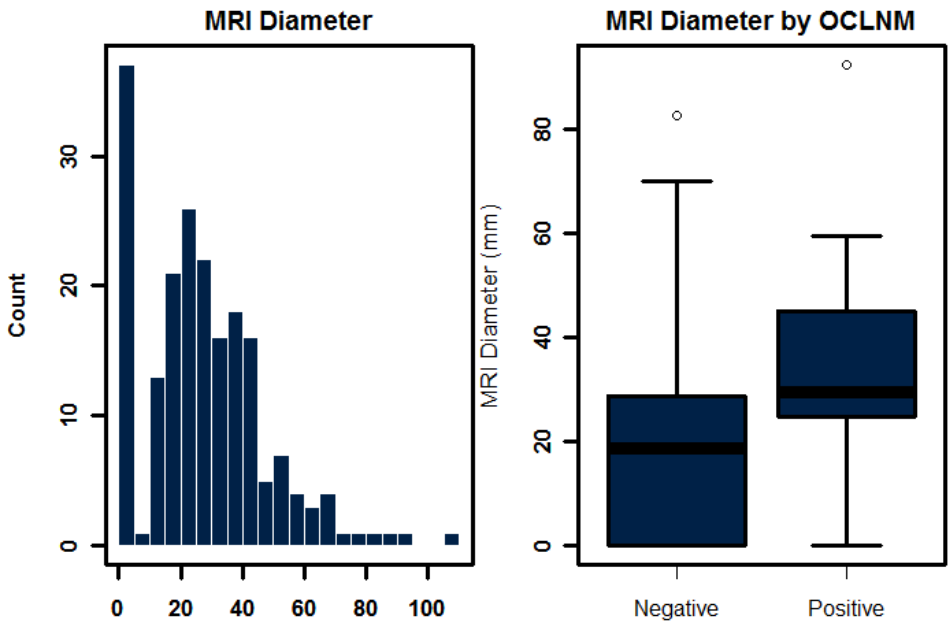


Figure 5.28: MRI diameter. On the left, a histogram of the distribution of MRI diameter. Each bar represents a five millimeter interval. In the right, a box plot of MRI diameter by OCLNM. The blue boxes represent the interquartile range (IQR) while the black line within the blue boxes represent the median. The whiskers extend a distance of 1.5 times the IQR from the upper and lower edges of the box. Extreme values are represented by hollow circles.

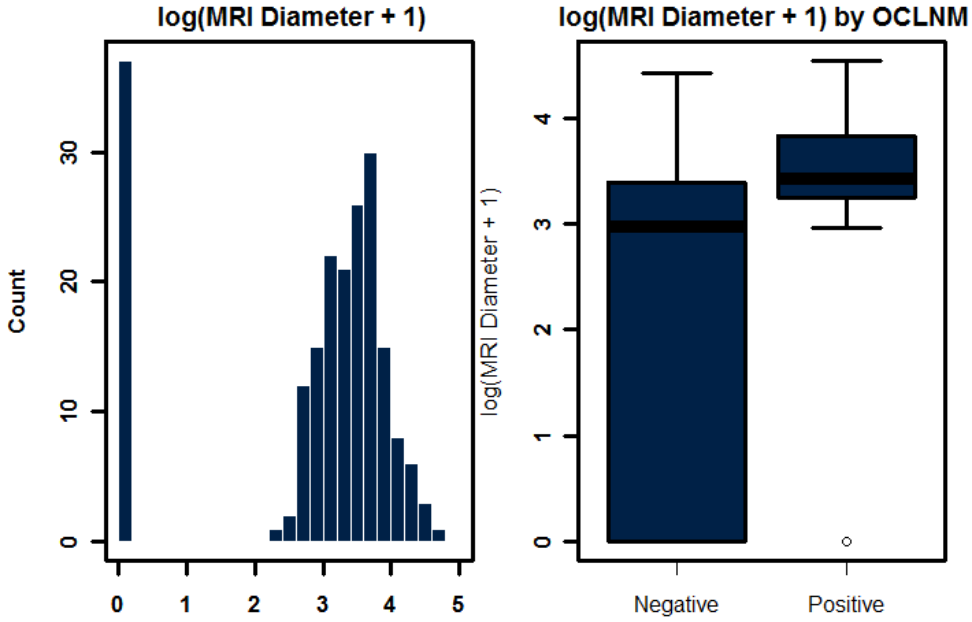


Figure 5.29: Log-transformed MRI diameter. On the left, a histogram of the distribution of log-transformed MRI diameter. Each bar represents a 0.2 unit interval. In the right, a box plot of log-transformed MRI diameter by OCLNM. The blue boxes represent the interquartile range (IQR) while the black line within the blue boxes represent the median. The whiskers extend a distance of 1.5 times the IQR from the upper and lower edges of the box. Extreme values are represented by hollow circles.

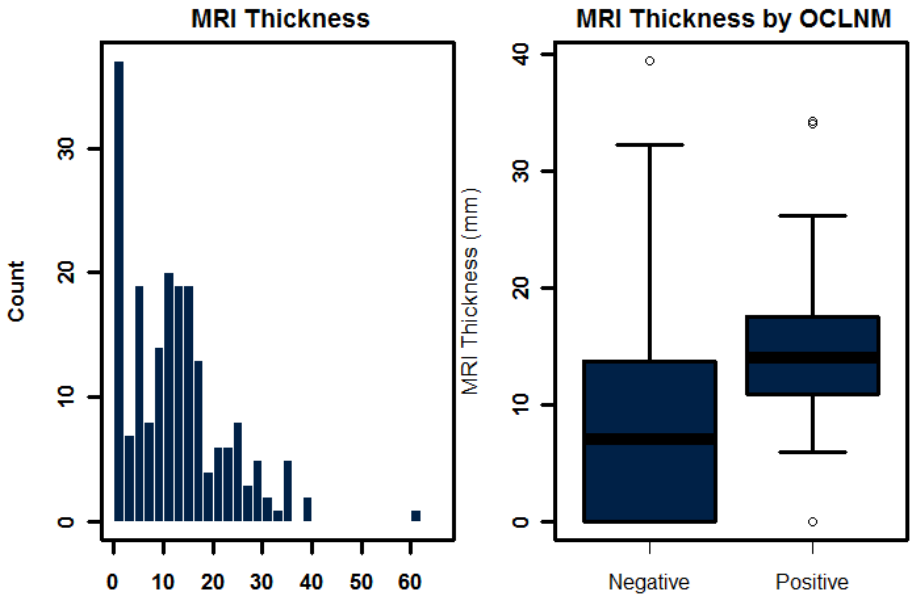


Figure 5.30: MRI thickness. On the left, a histogram of the distribution of MRI thickness. Each bar represents a 2 millimeter interval. In the right, a box plot of MRI thickness by OCLNM. The blue boxes represent the interquartile range (IQR) while the black line within the blue boxes represent the median. The whiskers extend a distance of 1.5 times the IQR from the upper and lower edges of the box. Extreme values are represented by hollow circles.

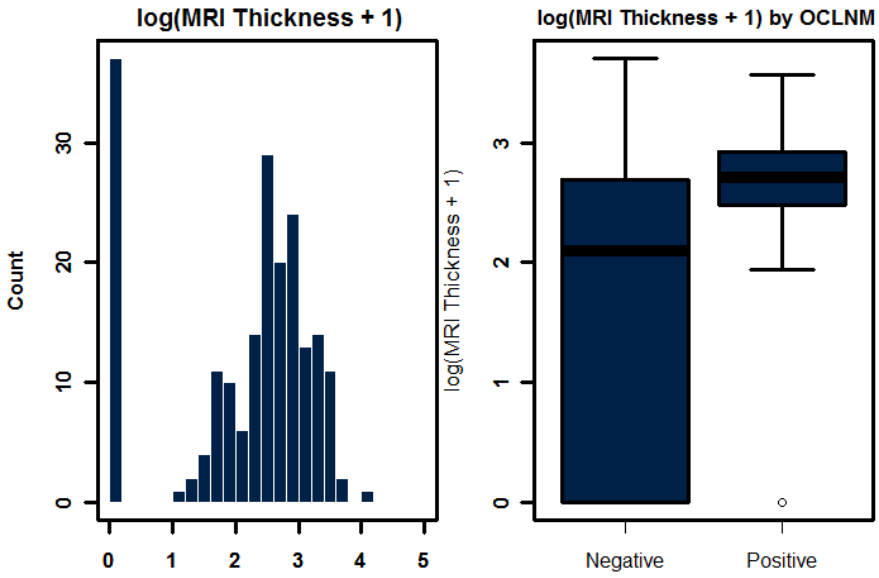


Figure 5.31: Log-transformed MRI thickness. On the left, a histogram of the distribution of log-transformed MRI thickness. Each bar represents a 0.2 unit interval. In the right, a box plot of log-transformed MRI thickness by OCLNM. The blue boxes represent the interquartile range (IQR) while the black line within the blue boxes represent the median. The whiskers extend a distance of 1.5 times the IQR from the upper and lower edges of the box. Extreme values are represented by hollow circles.

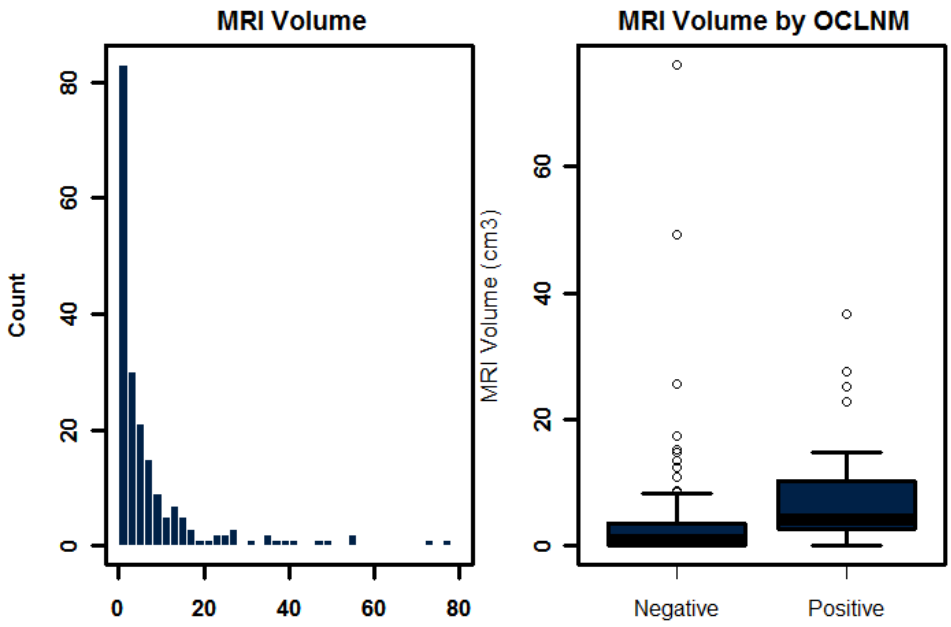


Figure 5.32: MRI volume. On the left, a histogram of the distribution of MRI volume. Each bar represents a 2 cubic centimeter interval. In the right, a box plot of MRI volume by OCLNM. The blue boxes represent the interquartile range (IQR) while the black line within the blue boxes represent the median. The whiskers extend a distance of 1.5 times the IQR from the upper and lower edges of the box. Extreme values are represented by hollow circles.

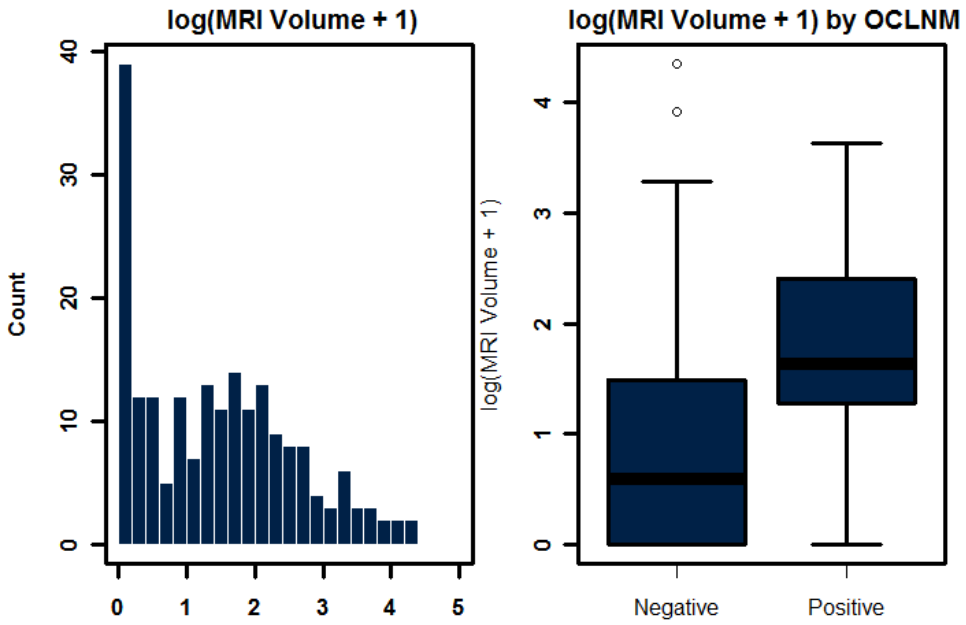


Figure 5.33: Log-transformed MRI volume. On the left, a histogram of the distribution of log-transformed MRI volume. Each bar represents a 0.2 unit interval. In the right, a box plot of log-transformed MRI volume by OCLNM. The blue boxes represent the interquartile range (IQR) while the black line within the blue boxes represent the median. The whiskers extend a distance of 1.5 times the IQR from the upper and lower edges of the box. Extreme values are represented by hollow circles.

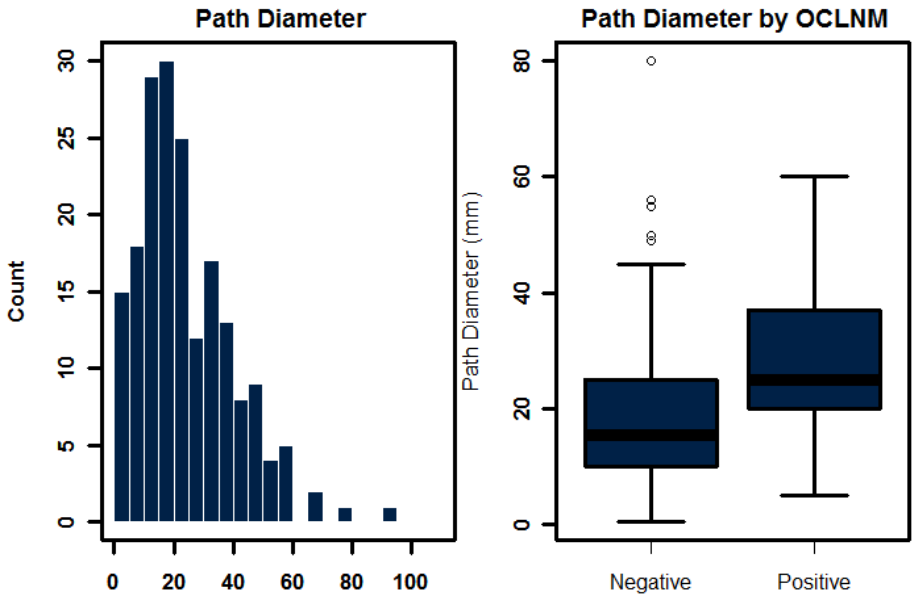


Figure 5.34: Pathology diameter. On the left, a histogram of the distribution of pathology diameter. Each bar represents a 5 mm interval. In the right, a box plot of pathology diameter by OCLNM. The blue boxes represent the interquartile range (IQR) while the black line within the blue boxes represent the median. The whiskers extend a distance of 1.5 times the IQR from the upper and lower edges of the box. Extreme values are represented by hollow circles.

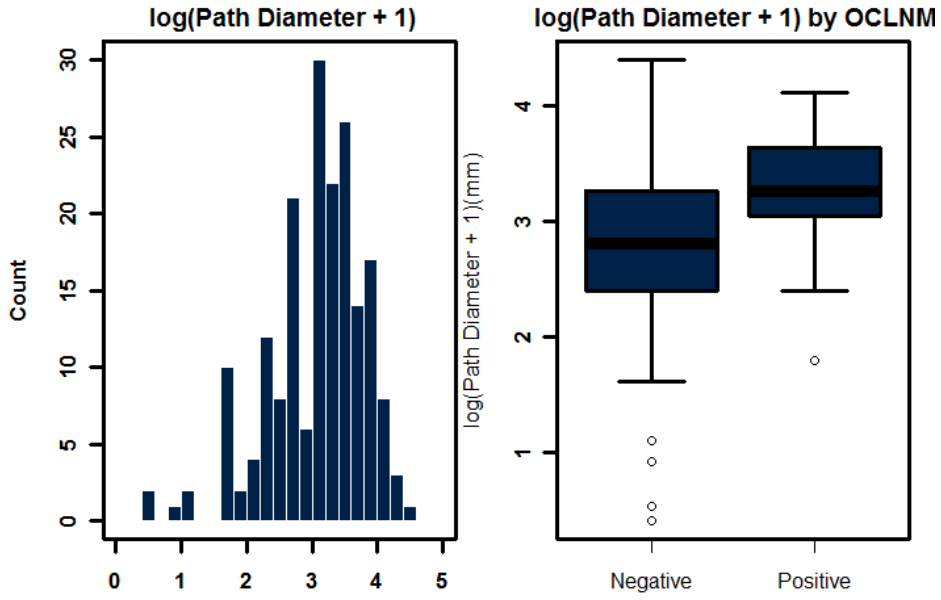


Figure 5.35: Log-transformed pathology diameter. On the left, a histogram of the distribution of log-transformed pathology diameter. Each bar represents a 0.2 unit interval. In the right, a box plot of log-transformed pathology diameter by OCLNM. The blue boxes represent the interquartile range (IQR) while the black line within the blue boxes represent the median. The whiskers extend a distance of 1.5 times the IQR from the upper and lower edges of the box. Extreme values are represented by hollow circles.

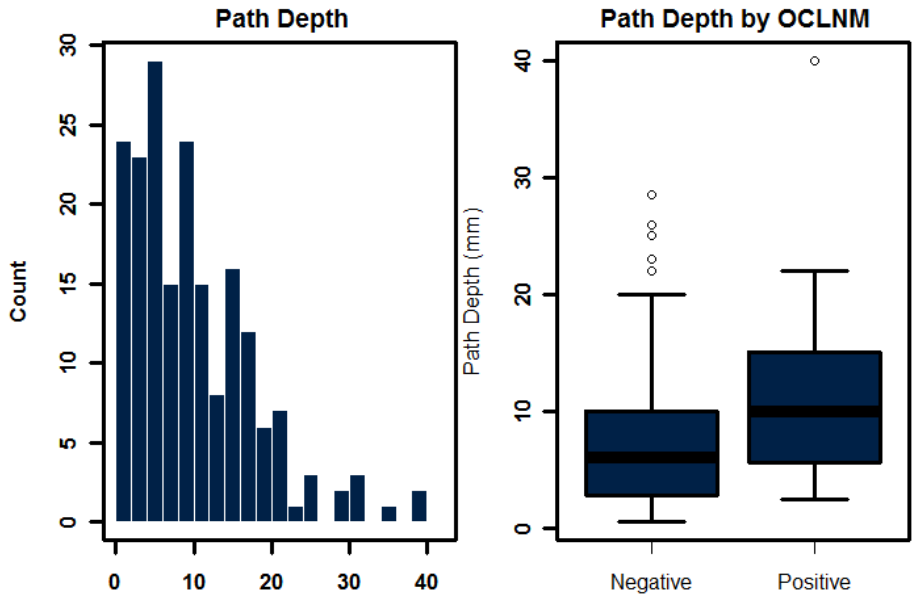


Figure 5.36: Pathology depth. On the left, a histogram of the distribution of pathology depth. Each bar represents a 2 mm interval. In the right, a box plot of pathology depth by OCLNM. The blue boxes represent the interquartile range (IQR) while the black line within the blue boxes represent the median. The whiskers extend a distance of 1.5 times the IQR from the upper and lower edges of the box. Extreme values are represented by hollow circles.

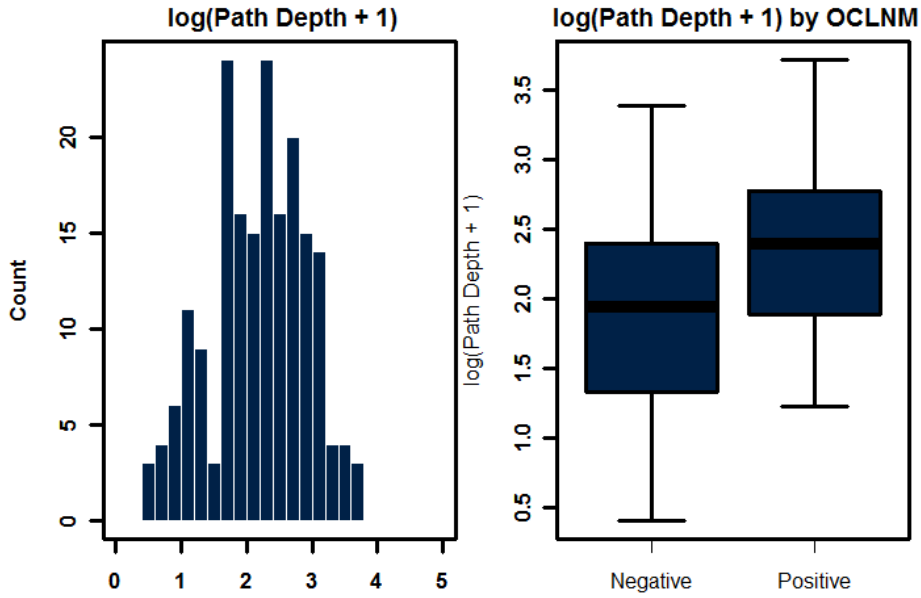


Figure 5.37: Log-transformed pathology depth. On the left, a histogram of the distribution of log-transformed pathology depth. Each bar represents a 0.2 unit interval. In the right, a box plot of log-transformed pathology depth by OCLNM. The blue boxes represent the interquartile range (IQR) while the black line within the blue boxes represent the median. The whiskers extend a distance of 1.5 times the IQR from the upper and lower edges of the box. Extreme values are represented by hollow circles.

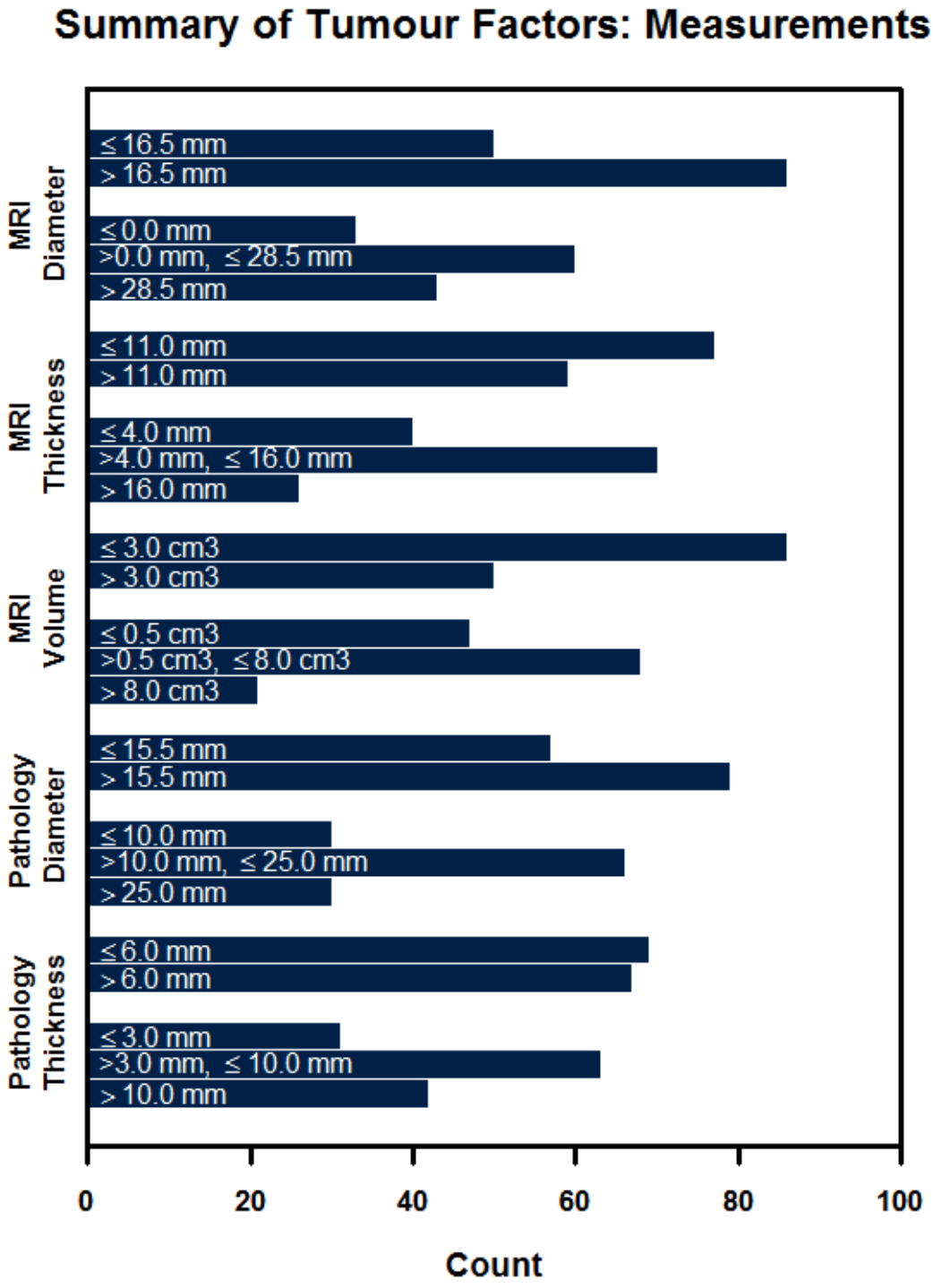


Figure 5.38: Distribution of categorized measurement tumour factors. MRI: magnetic resonance imaging.

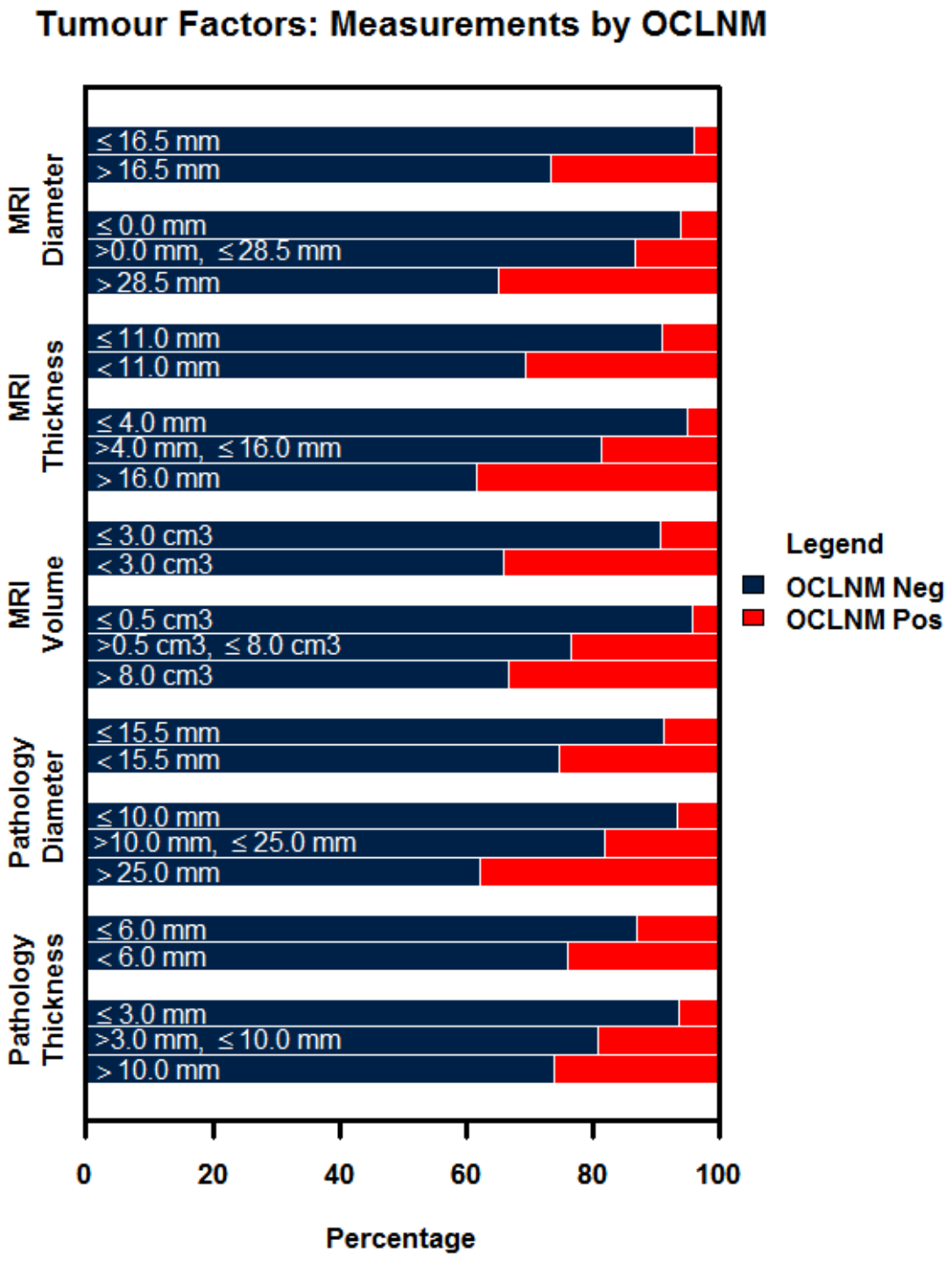


Figure 5.39: Categorized measurement tumour factors by OCLNM status. Proportion of cases negative for OCLNM (blue) and positive for OCLNM are shown (red). MRI: magnetic resonance imaging.

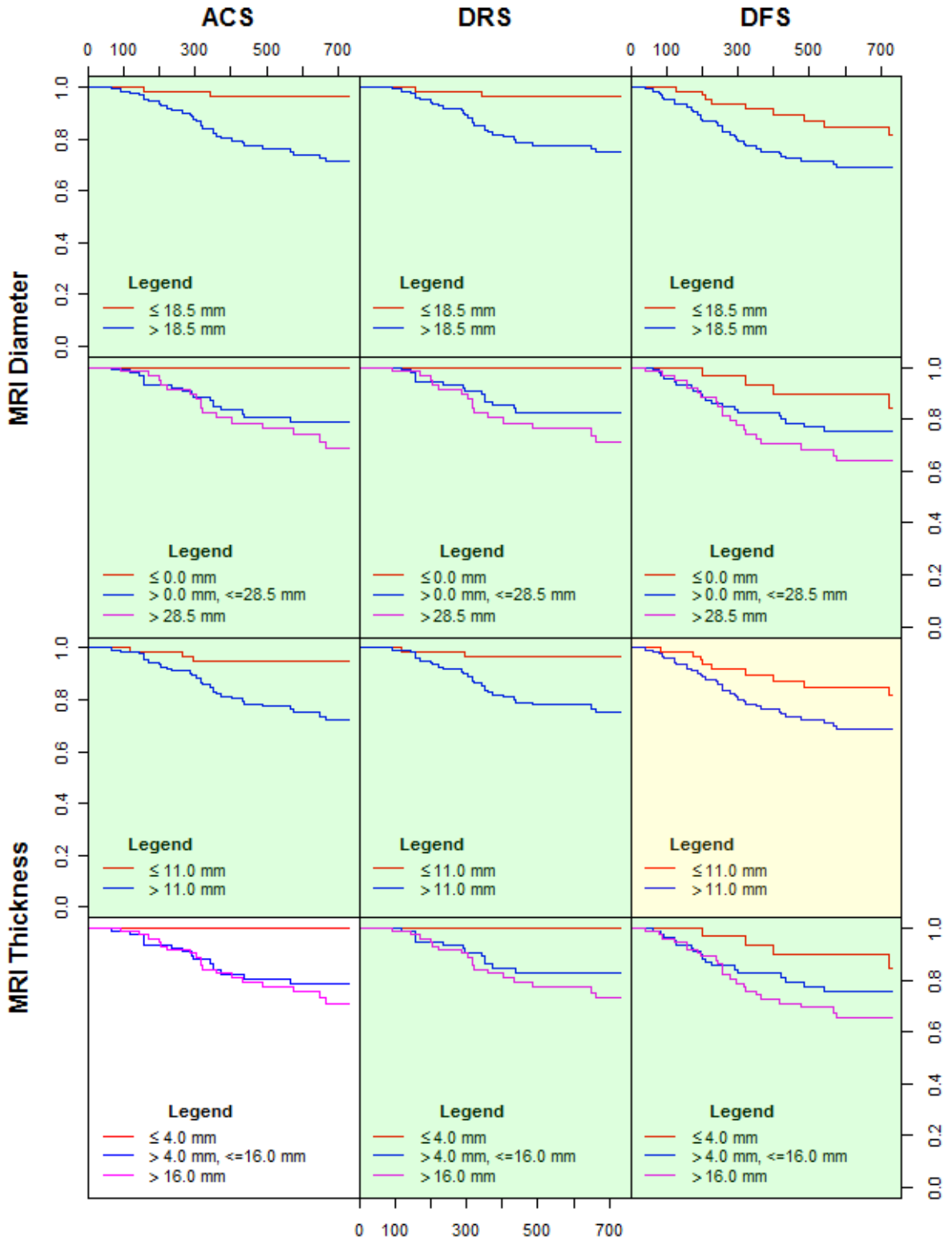


Figure 5.40: Survival by measurement tumour factors (1 of 3). Each square represents a Kaplan-Meier curve. The components involved each survival plot are indicated by the intersection of the measurement factors (left margin of grid) and the type of survival (top margin of grid). The x-axis scale of each plot is in days while the y-axis scale is cumulative survival fraction. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival; MRI: magnetic resonance imaging. Green indicates significant results. Yellow indicates trend towards significance (0.05 < p <= 0.10).

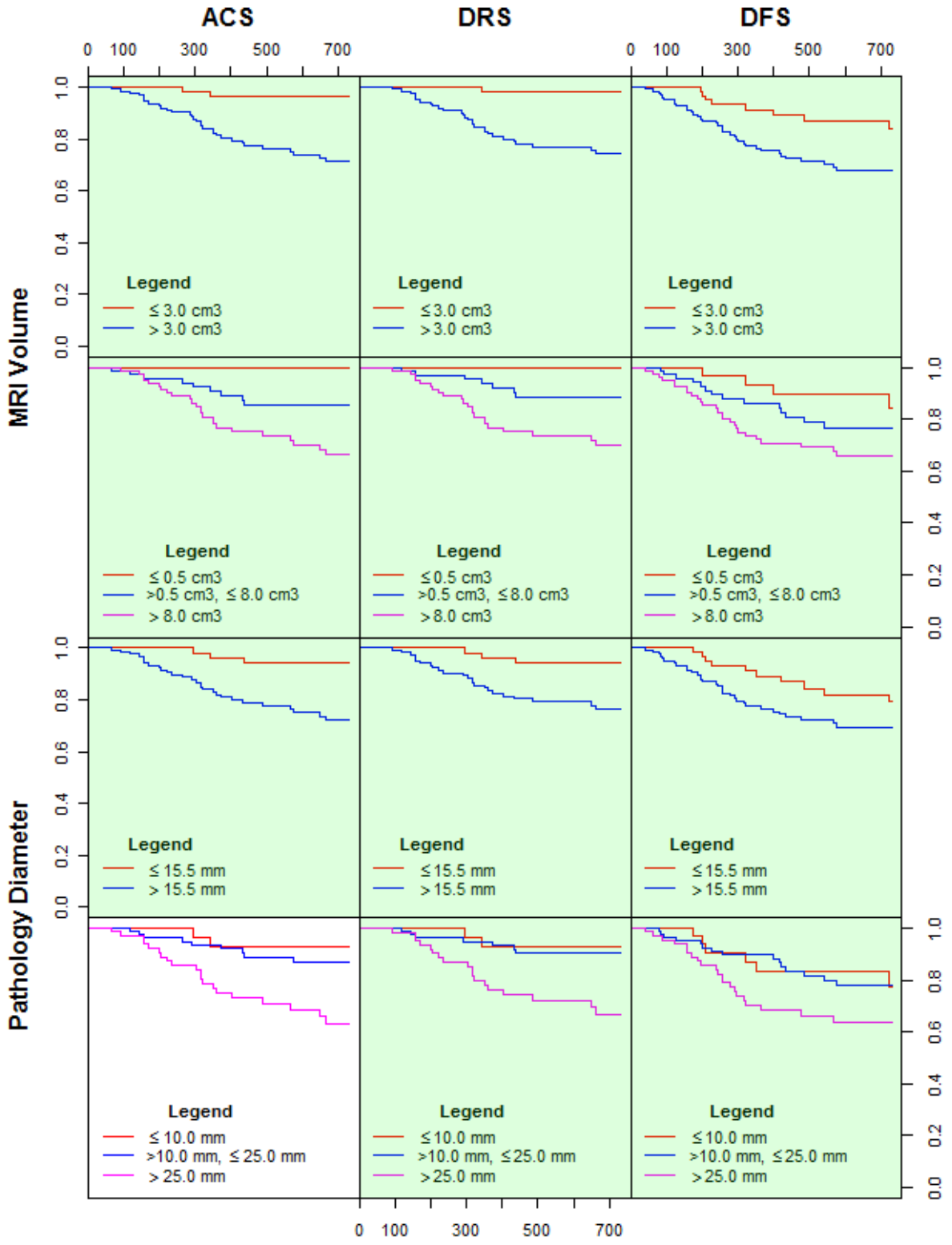


Figure 5.41: Survival by measurement tumour factors (2 of 3). Each square represents a Kaplan-Meier curve. The components involved each survival plot are indicated by the intersection of the measurement factors (left margin of grid) and the type of survival (top margin of grid). The x-axis scale of each plot is in days while the y-axis scale is cumulative survival fraction. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival; MRI: magnetic resonance imaging. Green indicates significant results.

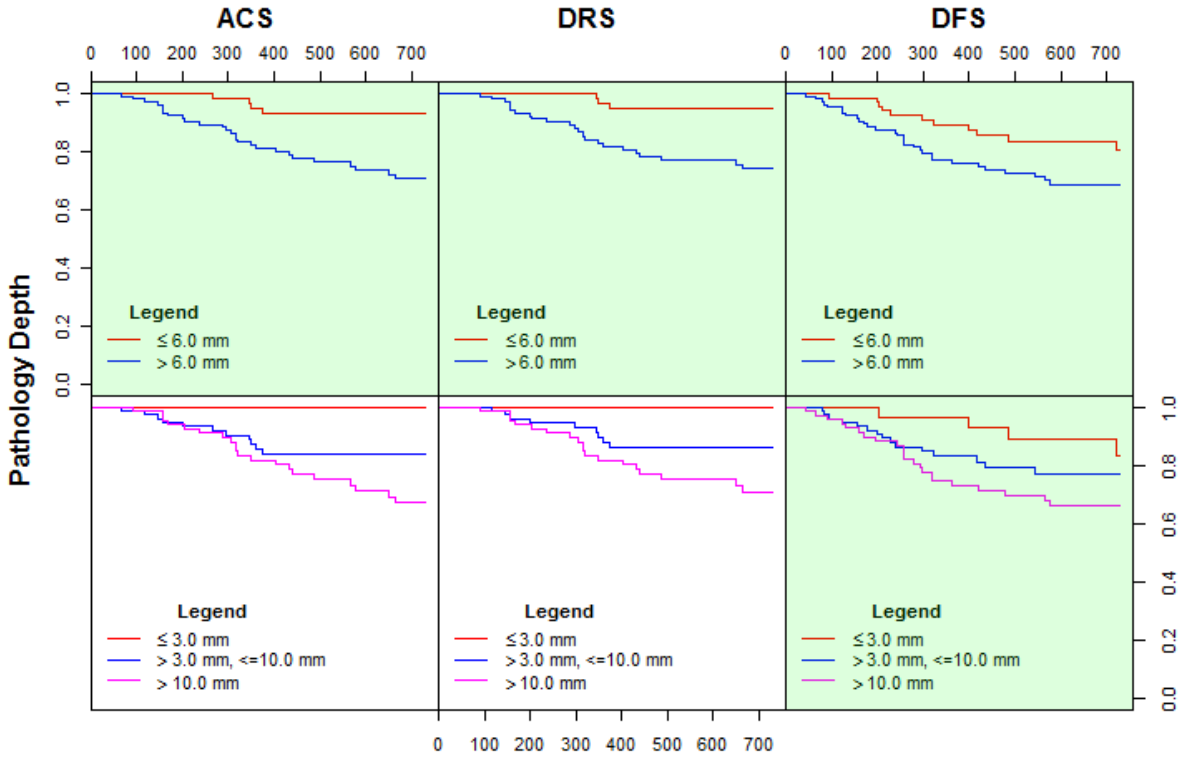


Figure 5.42: Survival by measurement tumour factors (3 of 3). Each square represents a Kaplan-Meier curve. The components involved each survival plot are indicated by the intersection of the measurement factors (left margin of grid) and the type of survival (top margin of grid). The x-axis scale of each plot is in days while the y-axis scale is cumulative survival fraction. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival; MRI: magnetic resonance imaging. Green indicates significant results.

5.9 Conclusions

This chapter presents a great deal of information about the ability of individual patient and tumour factors to predict occult cervical lymph node metastasis and survival. At first glance the information may seem overwhelming, but an overview of each individual factor is key to subsequent multivariate modeling in Chapter 6.

Based on the results of univariate analysis, a number of patient and tumour factors are available for inclusion in the multivariate modeling of OCLNM and survival. These factors include those with significant p-values (≤ 0.05) and those with a trend towards significance ($0.05 < p \leq 0.10$).

For pre-surgical prediction of OCLNM, the following factors will be carried forward for multivariate modeling:

MRI volume (continuous, log transformed and categorized)
MRI thickness (continuous, log transformed and categorized)
MRI diameter (continuous, log transformed and categorized)
Clinical T
Imaging bone invasion

For post-surgical prediction of OCLNM, the following factors will be carried forward for multivariate modeling:

MRI volume (continuous, log transformed and categorized)
Pathological depth (continuous, log transformed and categorized)
Pathological diameter (continuous, log transformed and categorized)
Pathological T
Pathological bone invasion
Tumour grade
Lymphovascular invasion
Perineural invasion

For pre-surgical prediction of survival, the following factors will be carried forward for multivariate modeling:

MRI volume (continuous, log transformed and categorized)
MRI thickness (continuous, log transformed and categorized)
MRI diameter (continuous, log transformed and categorized)
Clinical T
Clinical N
Clinical Stage
MRI extracapsular spread
Imaging bone invasion
Midline invasion
Age

For post-surgical prediction of survival, the following factors will be carried forward for multivariate modeling:

MRI volume (continuous, log transformed and categorized)
Pathological depth (continuous, log transformed and categorized)
Pathological diameter (continuous, log transformed and categorized)
Pathological T
Pathological N
Pathological Stage
Pathology extracapsular spread
Pathological bone invasion
Tumour grade
Invasive front
Lymphovascular invasion
Perineural invasion
Age

While clinical T-stage was predictive of OCLNM and survival, it remains a poor surrogate for pathological T-stage with cT and pT agreeing in only 60% of cases. There were few clinical or pathological T3 tumours, demonstrating that most tumours that fall into the greater than 40 mm in diameter category usually invade other key structures and are designated cT3. There remains a case for an alternative approach to TNM staging of oral cancers.

It is impossible to distinguish between MRI-based volume, diameter and thickness based on the results of univariate analysis. All seem to be of similar value in terms of OCLNM and survival

prediction. The most striking predictor of survival remains cervical lymph node metastasis. Both cN and pN were strong predictors of ACS, DRS and DFS with large hazard ratios, and will likely play a key role in any model produced.

The following chapter brings these individual factors together in a series of multivariate models designed to provide an individualized approach to the management of oral cancer. The reader is invited to refer back to this chapter for relevant discussion.

6 Multivariate analysis

6.1 Introduction

The previous chapter established the ability of individual patient and tumour-related factors to predict OCLNM and survival. In isolation, individual factors give insight into the natural history and prognosis of disease. No single factor is capable of reflecting the complexity of OCSCC. Multivariate models can optimize the prediction of prognosis and OCLNM by including multiple factors. However, model building is a complex process requiring a systematic approach.

There are a number of methods used to build multivariate models. Conventional statistical approaches include logistic regression for predicting binary outcomes (i.e. OCLNM) and Cox proportional hazards for survival. Artificial neural networks (ANNs) use "machine learning" to simulate biological learning processes and have the potential to learn complex, non-linear relationships. However, they are often referred to as "black box", making interpretation difficult. Recursive partitioning uses empirical data to construct decision trees for the outcome. The product is very intuitive to use, especially in medicine where decision making algorithms are commonplace.

Advanced multivariate modeling methods are poorly explored in the COSCC literature. The application of advanced methodologies, detailed in this chapter, represents novel work in the natural history and staging of oral cavity cancer.

This chapter presents detailed methodology for the construction of multivariate regression, ANN and recursive partition models using the factors established in Chapter 5, and the data base described in Chapter 3. The models constructed using this methodology are evaluated for their ability to predict OCLNM and survival, with emphasis on the capability to generalize to external data. Model-building methods are compared, and the best overall models are chosen for OCLNM and survival prediction.

6.1.1 Chapter goals

The goals of this chapter are:

1. *To provide the reader with a background in the theory of multivariate modeling, including*

logistic regression, Cox proportional hazards, ANNs and recursive partitioning.

2. *To outline how each of these modeling techniques can be applied to the present dataset for the prediction of OCLNM and survival.*

3. *To establish a rigorous methodology section that describes, in detail, how each of these techniques was used in this thesis.*

4. *To use the patient and tumour factors evaluated in Chapter 5 to construct regression, ANN and recursive partitioning models for OCLNM and survival.*

5. *To summarize and discuss the results of multivariate modeling for OCLNM and survival prediction.*

6. *To compare and contrast each model to establish which provides the most accurate method for predicting OCLNM and survival.*

7. *To provide a rigorous discussion of the implications of the results and how identified models can be used to improve the quality of information available to surgeons and patients about OCLNM probability and prognosis.*

8. *To describe how the established models can be used to provide information specific to each individual patient.*

6.1.2 Organization of sections

The chapter begins with an in-depth background section outlining the task of multivariate modeling, including theory specific to modeling binary outcomes and censored survival data. Each individual modeling technique is discussed, including logistic regression, Cox proportional hazards, ANNs and recursive partitioning. Advantages and disadvantages are given. This is followed by a brief section outlining how models are evaluated, including the ability to generalize and error measurement.

The methodology section is extensive and begins with an overview of explanatory variables to be used in modeling, as well as a general modeling procedure. The methodology specific to each individual technique is then discussed for both OCLNM and survival prediction. Methods for measuring model error and comparing individual models are outlined.

The results and discussion sections depart from the traditional linear format. Instead, there are two

separate groupings of results and discussion for OCLNM and survival prediction. Within each section, the results for pre-surgical and post-surgical prediction are given separately, followed by a unifying discussion. It was felt that this compartmentalized organization lends structure to an otherwise large amount of information that might prove confusing otherwise.

This is followed by a general discussion that brings together common themes. Time is spent discussing the potential shortcomings of multivariate modeling that must be considered when interpreting the results. The chapter ends with a summary and conclusions.

6.2 Background

This section presents methodology specific to this Chapter. While the background is extensive, it is necessary in order to understand the methods and results outlined later in the Chapter. A overview of classification is given, followed by background on logistic regression, ANNs and recursive partitioning. This is followed by an overview of censored survival data theory. Cox proportional hazards and adaptation of ANNs and recursive partitioning to survival data are discussed.

6.2.1 Classification

The ability of any classification rule to predict a binary outcome is depicted in Table 6.1 [206].

Table 6.1: Test results versus actual outcome.

		Outcome		
		Positive	*Negative*	
Test	*Positive*	*a*	*b*	*a+b*
	Negative	*c*	*d*	*c+d*
		a+c	*b+d*	*a+b+c+d*

where "Test" refers to the predicted result and "outcome" refers to the observed result. In this format, four possible situations are specified:

"a" - True positive results in which the test correctly predicts a positive outcome.

"b" - False positive results in which the test incorrectly predicts a positive outcome.

"c" - False negative results in which the test incorrectly predicts a negative outcome.

"d" - True negative results in which the test correctly predicts a negative outcome.

From these situations it is possible to define five terms common in medical literature involving binary outcomes. Knowledge of these terms is necessary to assess the value of predictive models [206].

Equation 6-1

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives} = \frac{a}{a + c}$$

Equation 6-2

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives} = \frac{d}{b + d}$$

Equation 6-3

$$Positive\ Predictive\ Value = \frac{True\ Positives}{True\ Positives + False\ Positives} = \frac{a}{a + b}$$

Equation 6-4

$$Negative\ Predictive\ Value = \frac{True\ Negatives}{True\ Negatives + False\ Negatives} = \frac{d}{c + d}$$

Equation 6-5

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Cases} = \frac{a + d}{a + b + c + d}$$

The predictive utility of a test is often measured in terms of accuracy. However, in many situations where one outcome has a particularly detrimental effect, it may be desirable to minimize false negative results. One such example in head and neck oncology is OCLNM. Because CLNM is associated with a 50% decrease in disease-related survival [3], the negative predictive value (NPV) of any test must be high at the expense of positive predictive value (PPV) and accuracy.

6.2.2 Logistic regression models for classification

When a binary outcome is involved, logistic regression can be used to assess the value of one or

more explanatory variables. These explanatory variables may be categorical or continuous.

6.2.2.1 Theory

The following derivation is adapted from Ostir and Uchida [207]. Consider a binary outcome variable, *y*, with values coded as 0 and 1, and an explanatory variable, *x*. If the probability of the outcome *y* varies according to the sigmoidal curve as a function of *x*, then the probability that *y = 1*, *π*, is expressed as a logistic regression function:

Equation 6-6

$$Probability\ [y = 1] = \pi = \frac{\exp\ (\propto + \beta x)}{1 + \exp\ (\propto + \beta x)}$$

where "*exp*" refers to the exponential function and the coefficients *α* and *β* define the relationship between *y* and *x*. The logit transformation is used to change the form of the independent variable, *π*, to the natural logarithm of *π/(π-1)*.

Equation 6-7

$$\ln \frac{\pi}{1 - \pi} = \ln \left[\left(\frac{\exp\ (\propto + \beta x)}{1 + \exp\ (\propto + \beta x)} \right) \Big/ \left(1 - \frac{\exp\ (\propto + \beta x)}{1 + \exp\ (\propto + \beta x)} \right) \right]$$

This reduces to:

Equation 6-8

$$\ln \frac{\pi}{1 - \pi} = \ln[\exp\ (\propto + \beta x)] = \propto + \beta x$$

The odds that the outcome variable, *y*, is equal to *1* is the ratio of the probability of *y = 1*, *π*, to the probability of *y ≠ 1*, *1 - π*.

Equation 6-9

$$Odds\ [y = 1] = \frac{\pi}{1 - \pi}$$

The log odds are defined by taking the natural logarithm of the previous equation.

Equation 6-10

$$\ln(Odds\ [y = 1])\ = \ln\left(\frac{\pi}{1-\pi}\right)$$

Therefore:

Equation 6-11

$$\ln(Odds\ [y = 1])\ = \propto + \beta x$$

and

Equation 6-12

$$Odds\ [y = 1] = exp(\propto + \beta x)$$

This equation can be used to express the odds ratio (OR), a parameter frequently quoted in the medical literature. The OR is a way of expressing the odds of the outcome *(y = 1)* at a given value of the explanatory variable, x_i, compared to another value of the explanatory variable, x_j.

Equation 6-13

$$Odds\ Ratio\ (OR) = \frac{odds\ [y = 1\ at\ x = x_j]}{odds\ [y = 1\ at\ x = x_i]}$$

Equation 6-14

$$Odds\ Ratio\ (OR) = \frac{exp(\propto + \beta x_j)}{exp(\propto + \beta x_i)}$$

Equation 6-15

$$Odds\ Ratio\ (OR)\ = exp\{(\propto + \beta x_j) - (\propto + \beta x_i)\}$$

Equation 6-16

$$Odds\ Ratio\ (OR)\ = exp\{\beta(x_j - x_i)\}$$

Multiple logistic regression is an extension of logistic regression when two or more explanatory variables are used. The multiple logistic regression equation using *p* explanatory variables can be expressed as:

Equation 6-17

$$Probability\ [y = 1] = \pi = \frac{exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}{1 + exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}$$

Similarly the multiple explanatory variable odds and OR can be expressed as:

Equation 6-18

$$Odds\ [y = 1] = \frac{\pi}{1 - \pi} = exp\left(exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)\right)$$

and

Equation 6-19

$$Odds\ Ratio\ (OR) = exp\{\beta_1(x_{1j} - x_{1j}) + \beta_2(x_{2j} - x_{2j}) + \cdots + \beta_p(x_{pj} - x_{pj})\}.$$

6.2.2.2 Advantages

The characteristics of logistic regression make it useful in the medical literature. Medical research often involves binary outcomes, such as the presence or absence of disease [207-208]. These outcomes are often the result of complex processes that have multiple explanatory variables [208]. Logistic regression is flexible enough to handle a heterogeneous group of explanatory variables, which may be continuous, categorical or a combination of both [207-208].

Using logistic regression, the relationships between outcome and explanatory factors are expressed as odds ratios, a measure of the magnitude of effect that is common in the medical literature and easy to interpret [207-208]. ORs allow comparison of the relative contribution of each factor to the final model [207-208]. Further, the final logistic regression equation allows clinicians to calculate an individualized probability for the outcome based on the patient's explanatory variables [208].

6.2.2.3 Disadvantages

Logistic regression, though widely used, is not without disadvantages. In some situations,

interactions can exist between variables. Knowledge of such interactions can improve accuracy. These interactions seldom obvious and, when a large number of explanatory variables are involved, investigation of all possible higher order interactions is difficult [208].

Logistic regression also assumes that there is a linear relationship between explanatory variables and the outcome. These non-linear relationship must be investigated and resolved with various transformations, but the work is time and resource consuming. Other methods, like ANNs and recursive partitioning can cope with non-linearity and provide more accurate results [208].

6.2.3 Artificial neural networks (ANNs) for classification

ANNs are computational models that mimic the information flow in the human brain. Predictive factors can be thought of as network input, which is then used to formulate a prediction, or network output. Between the input and output “layers” are one or more “hidden layers” that form interconnections that parallel neuronal complexity.

6.2.3.1 Theory

ANNs are composed of individual units known as processing elements (PEs). Each PE is composed of inputs, weighted connections, an activation function, a transfer function and a single output [209]. Each arriving input, *X*, is multiplied by a connection weight, *W*. The weighted inputs are summed and if the sum exceeds the "activation" threshold, the neuron's output is determined by the transfer function [210]. The most commonly used transfer function is the sigmoid curve. Alternatives include linear, tansig and softmax [209]. A single neuron is shown in Figure 6.1.

The power of ANNs stem from the interconnection of individual PEs to form a complex network. Much like the human brain, the functional value of the whole exceeds the sum of its constituent parts. The most common ANN structure used in the medical literature consists of a layer of input neurons, a single layer of intermediate neurons, called "hidden neurons", and a layer of output neurons [210]. An entire network is shown in Figure 6.2.

Once the structure is established, the network is supplied with "training data" consisting of inputs and known outputs. The network feeds the input data through the first series of connection weights to the hidden layer, then through the second set of connection weights to the output layer where the final output is determined. This structure is often referred to as a "feed-forward" ANN [210].

The connection weights are central to the ANN's ability to "learn" and recognize patterns in the input data that allows it to predict the outcome. Like biological neurons, these weights can determine if a neuron is excitatory or inhibitory, or has a large or small influence [209]. Initially, the weights are set randomly and the output has a high prediction error. However, this is back-propagated to the network and the connecting weights adjusted to improve the prediction error. This process occurs iteratively until the prediction error has been minimized [210]. Once performance has been optimized using the training data, the ANN is capable of making predictions when presented with novel input data.

6.2.3.2 Advantages

Unlike logistic regression, ANNs are not limited by the assumption that relationships between explanatory variables and the outcome are linear. ANNs are capable of adapting to non-linearity by optimizing the weighted interconnections between neurons [208]. Theoretically, this saves researchers from having to search and account for non-linearity. Further, the network can integrate new input cases in a process of continuous training that allows the network to evolve over time.

6.2.3.3 Disadvantages

The major disadvantage of the ANN is the “black box” nature of the technique. While the method can model complex data, there is no intuitive way to define the relative importance of each variable [208].

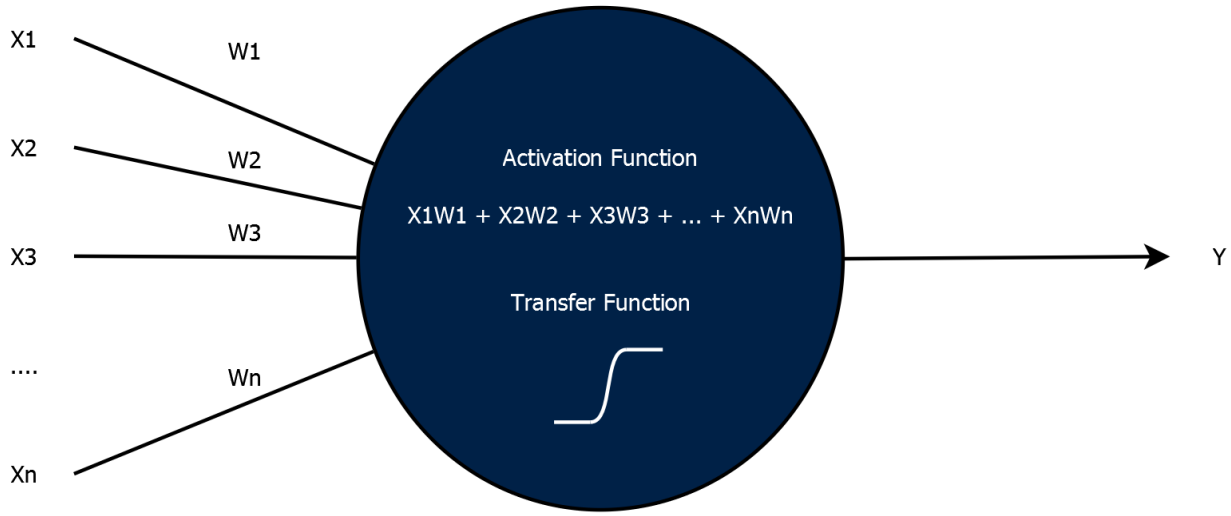


Figure 6.1: A neural network processing element. Inputs, *X*, are multiplied by a connection weight, *W*. The sum of the weighted inputs are passed to an activation function and the neuron output determined by the transfer function.

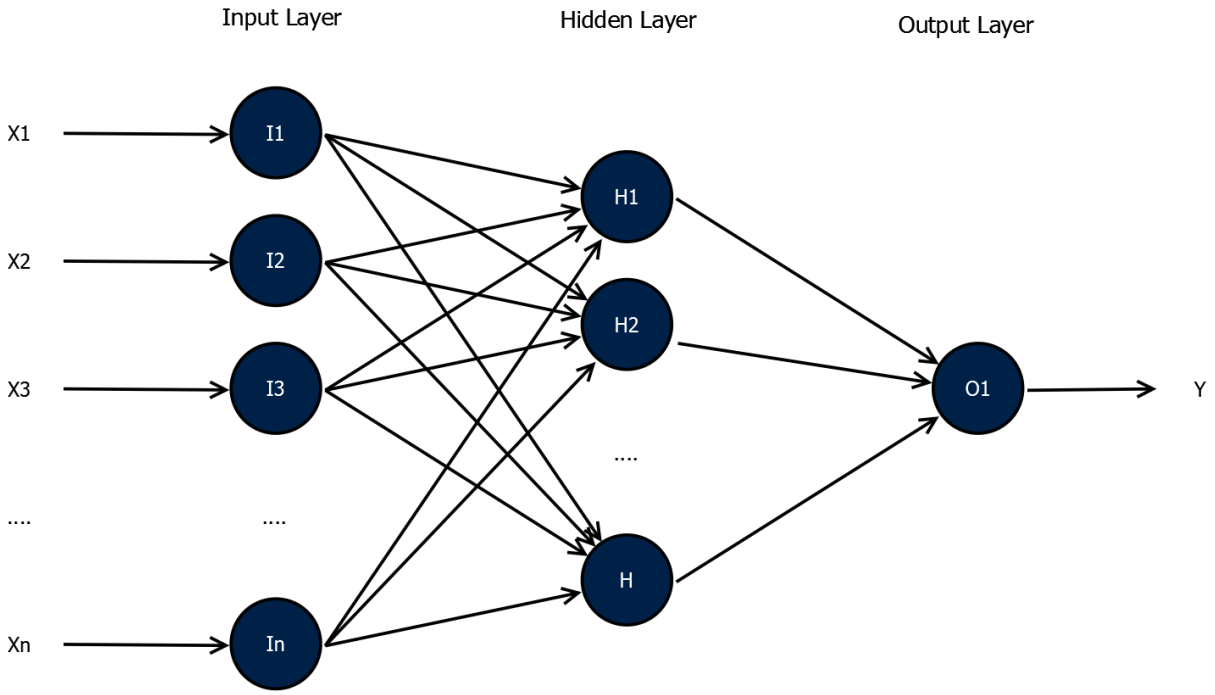


Figure 6.2: Feed-forward artificial neural network. Inputs, x, are passed to input neurons, *I*. The inputs are passed to a single layer of hidden neurons, *H*. In this structure a hidden output node, *O*, provides the network output, *Y*.

This in contrast to logistic regression in which the OR of each variable is obtained from the final model. This is a barrier to clinical dissemination making ANNs less likely to be adopted widely [208].

The advantages of ANN development do not necessarily translate to improved prediction. While many authors have shown that ANNs provide superior prediction in some situations, other authors have shown that ANNs are inferior to logistic regression techniques [211]. ANN training and development is computationally intensive. Large numbers of training inputs, as well an increased network complexity, may require excessive computational resources [210].

6.2.4 Recursive partitioning for classification

Recursive partitioning reflects the decision making processes that occur on a daily basis in the clinical environment. The basic premise is that a sample is progressively split into smaller sub-groups based on the predictive ability of explanatory variables.

6.2.4.1 Theory

The basic unit of a recursive partitioning decision tree, shown in Figure 6.3, consists of a parent node which is partitioned into two child nodes based on a splitting rule. In a full decision tree, this splitting is recursive in that child nodes become the parent node and are themselves split into two child nodes until the tree reaches a pre-determined level of complexity. In the full tree, the highest level node is referred to as the root node, while unsplit child nodes are referred to as terminal nodes [208].

The first step in recursive partitioning is the initiation of tree building. Building begins at the root node where a computer algorithm tests a series of splitting rules based on every possible value of each explanatory variable. The "best split" is determined to be the candidate split that minimizes the impurity of the two child nodes [210]. The impurity is determined using the "Gini coefficient" [212].

Equation 6-20

$$Gini\ Coefficient = i(p) = 1 - \sum_{l} p_l^2$$

where p refers to the relative frequencies of each class of the outcome variable (p_1, p_2, p_3, ..., p_l). The impurity measurement is weighted according to the proportion of parent node cases in each child node such that the reduction in impurity is given by:

Equation 6-21

$$Impurity\ Reduction = i(p)_{root} - \left(\pi_A i(p_A) + \pi_B i(p_B)\right)$$

where $i(p)_{root}$, $i(p)_A$ and $i(p)_B$ refer to the impurity of the root node and each child node respectively, and π_A and π_B refer to the proportion of each class of the outcome variable in each child node. The "best split" selection algorithm is then iteratively carried out for each child node until tree building is stopped.

The second step of tree building consists of stopping the iterative splitting process. Stopping occurs when any of a series of parameters is encountered [210].

> *The number of observations in a parent node is below a user-defined threshold value.*
> *The number of observations in a potential child node is below a user-defined threshold.*
> *All of the observations within a parent node are of a single class and thus "pure".*
> *The number of levels in the decision tree, numbered beginning from 0 at the root node, reaches a user-defined value.*

The result of the stopping process is a series of terminal nodes. Each terminal node is associated with a probability for the outcome of interest.

The final step of tree building is tree-pruning and optimal tree selection. Pruning refers to the removal of nodes from the stopped tree beginning with the most distant branches. The degree of pruning that takes place depends on a user-defined complexity parameter (CP). The CP is a measure of the additional reduction in impurity that a splitting criterion must offer in order to be included in the final tree structure. As the CP is increased, the pruning algorithm removes increasingly important splits from the tree, resulting in a less complex structure [210].

Tree stopping and pruning are necessary to avoid over-fitting of the decision tree. Techniques to measure the ability to generalize are discussed in Section 6.2.9. A representative example of a decision tree is shown in Figure 6.4.

6.2.4.2 Advantages

The main advantage of recursive partitioning is the intuitive nature of the resulting model. The decision tree is so named because it reflects structured clinical decision making. The difference is that the recursive partitioning algorithm is based on empirical data. In clinical environments, the decision tree model can be easily applied to known explanatory variables. This is compared to the black box output of ANNs, and the complex equations of logistic regression [208].

An additional advantage that recursive partitioning shares with ANNs is the ability to model complex interactions between explanatory variables and outcomes. Recursive partitioning does not require that assumptions be made with respect to linearity [208].

6.2.4.3 Disadvantages

It is tempting to conclude that the explanatory variables selected by recursive partitioning are the most important variables for determining the outcome. This might not be the case. Other authors have noted that small changes in the training data set can lead to variations in the splitting rules chosen [210]. Similarly, the final tree is not guaranteed to be the best. In some cases, a sub-optimal split early in the tree can lead to a better final tree structure. However, the Gini coefficient method means that the optimal split is chosen before moving on to the next tree level. As a result, there is no tree-construction mechanism that emphasizes future splitting decisions [210]. Caution is therefore required when using recursive partitioning to evaluate the relative importance of explanatory variables. This situation is further complicated when interactions exist amongst explanatory variables. A comprehensive evaluation of the explanatory variables requires study of not only the optimal splits but also the surrogate splits, which are candidate splitting criteria that were not included in the final model [210].

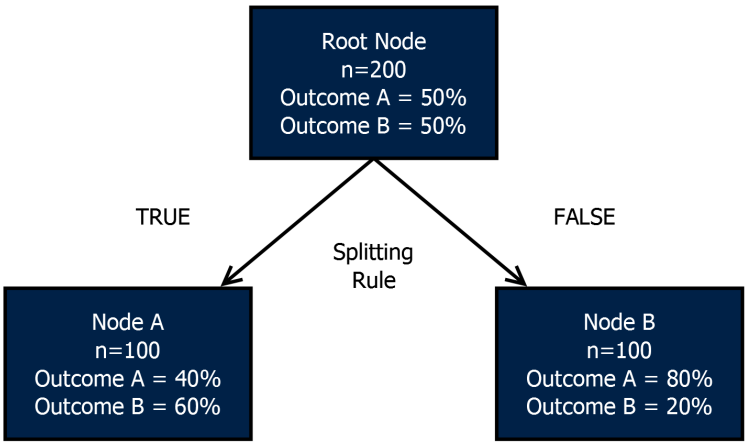


Figure 6.3: A example of a single split in a decision tree. The parent node is partitioned into two child nodes based on a binary splitting rule. Splitting is based on an "impurity" measure.

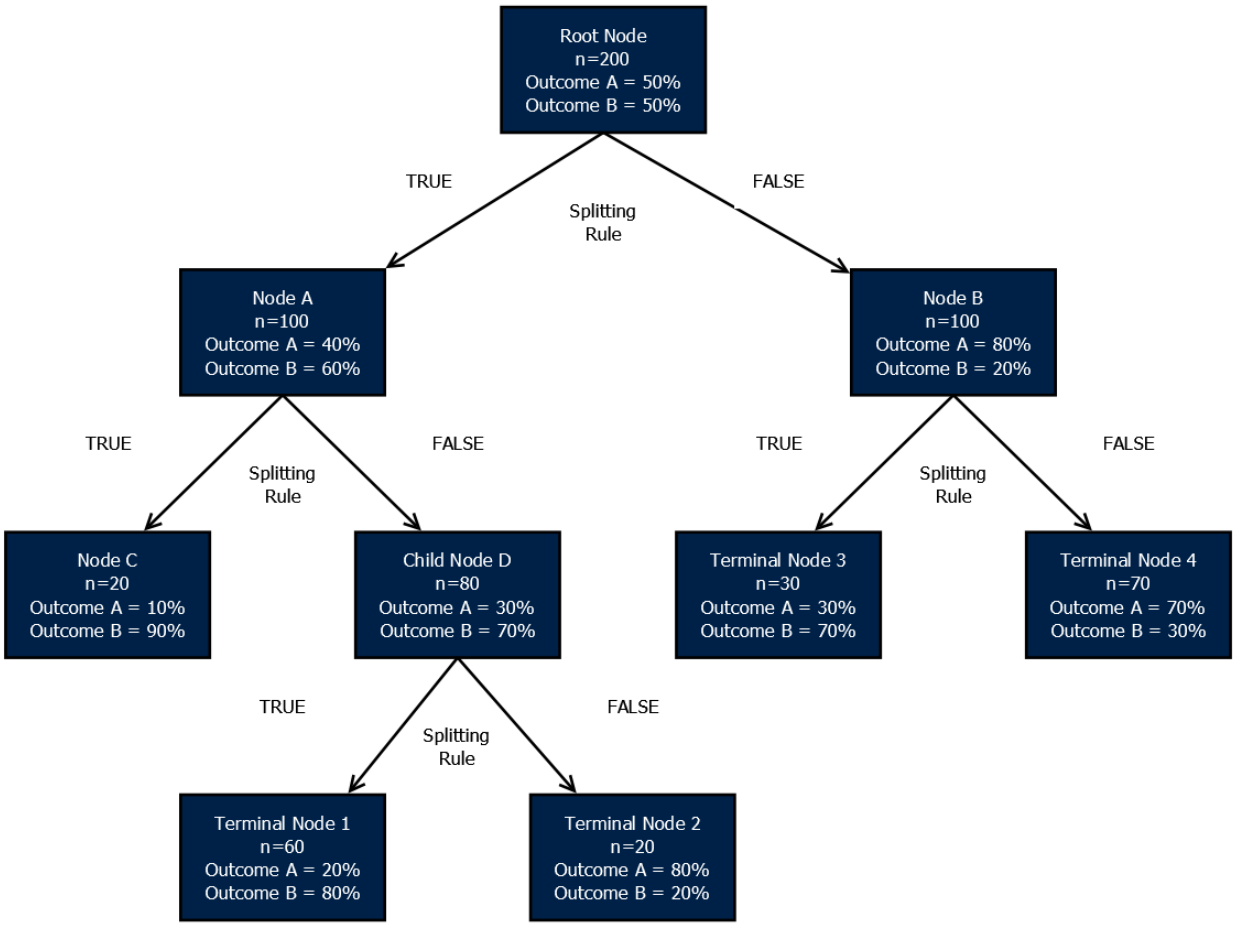


Figure 6.4: A complete decision tree. The example given shows a three-level tree with four splitting rules and five terminal nodes.

Recursive partitioning can be computationally intense. At each node, the recursive partitioning algorithm tests every possible splitting criterion. When the number of variables is small, the number of observations is large or the tree structure complex, considerable processing power is required.

6.2.5 Censored survival data

In survival analysis, patients may not experience the event of interest during follow-up period. These patients are said to be censored. There are several types of censoring encountered in studies of disease prognosis, diagrammed in Figure 6.5.

Type I censoring occurs when a patient survives for the duration of the study period without experiencing the event. For example, a patient in a study of 5-year DRS survives for 5 years [213]. Type II censoring is occurs when a patient is lost to follow-up prior to the end of the study period and prior to experiencing an event. An example is a patient who moves to a new location under the care of a different healthcare service [213]. Type III censoring is a combination of type I and II censoring. Types I, II and III censoring are also referred to as right censoring. This is compared to left censoring, which occurs when the time of disease acquisition is unknown [213].

For a group of right-censored individuals, Kaplan-Meier (KM) actuarial life tables and product-limit estimators can be used to estimate survival. These models assume that survival at time *t* is dependent on an individual's survival status at time *t - 1* [206]. The probability of survival at time *t* is estimated using the survivor function:

Equation 6-22

$$S(t) = \prod \left(\frac{n_j - d_j}{n_j} \right)$$

where n_j is the number of individuals at risk in time interval *j*, d_j is the number of deaths in the time interval and the product is for all intervals from 1 to *j*. Life tables and product-limit estimators differ in how the interval, *j*, is defined. For life tables, intervals of equal length are defined while for product-limit estimators intervals of variable length are defined for each event [206].

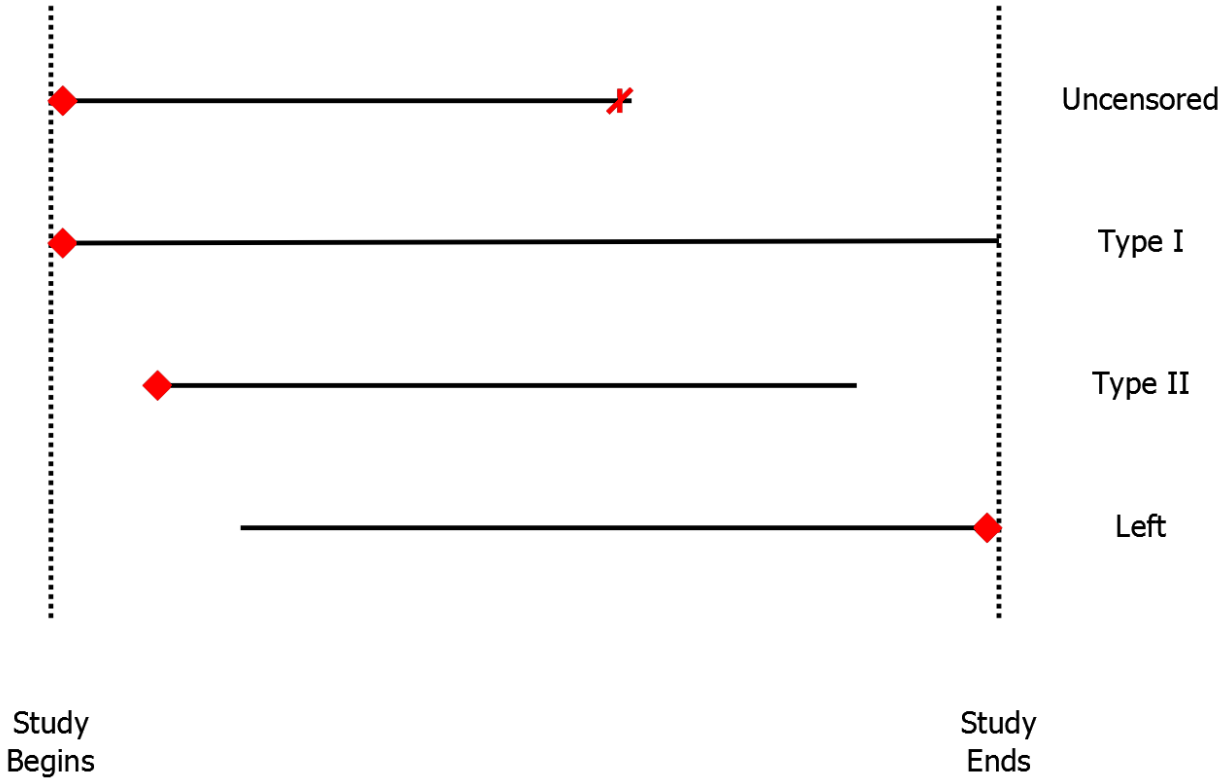


Figure 6.5: Types of censoring. The top line represents a case in which an event has occurred (*X*). Type one censoring occurs at the study endpoint. Type two censoring occurs when the case is lost to follow-up. Left censoring occurs when the time of entry into the study is unknown.

6.2.6 Cox proportional hazards for censored survival data

The Cox proportional hazards model is a semi-parametric model that is commonly used in medical studies of censored survival data [213]. The popularity of the Cox model stems from the fact that it is able to examine each explanatory variable's relationship with the outcome. This relationship is expressed as a hazard ratio (HR), a value familiar to the medical community. In this manner, the most important variables can be identified. Similarly, the Cox model is able to control for confounding factors when necessary [213].

The Cox model is based on the concept of proportional hazards, shown in the equation:

Equation 6-23

$$h_i(t) = h_0(t)e^{x_i\beta}$$

where $h_i(t)$ is the hazard for an individual *i* at time *t*, *x* is the vector of explanatory variables and β is the vector of coefficients associated with each explanatory variable [213]. Conceptually, the Cox model assumes that there is a baseline hazard common to all individuals. The hazard for any given individual is a multiple of that baseline, based on the value of the explanatory variables. All hazards are "proportional" to one another [214].

The proportional hazards assumption is central to calculating the HR. If we consider an explanatory variable with values x_1 and x_2, the hazard ratio can be expressed as:

Equation 6-24

$$Hazard\ Ratio\ (HR) = \frac{h_{2i}(t)}{h_{1i}(t)} = \frac{h_0(t)e^{x_2\beta}}{h_0(t)e^{x_1\beta}}$$

Equation 6-25

$$Hazard\ Ratio\ (HR) = \frac{e^{x_2\beta}}{e^{x_1\beta}}$$

Equation 6-26

$$Hazard\ Ratio\ (HR) = e^{\beta(x_2 - x_1)}$$

The resulting HR is independent of the baseline hazard. Though the baseline hazard need not be known for calculation of the HR, it is required for the calculation of survival probabilities for individual cases. Various methods can be used to estimate survival from Cox models [214]. However a complete discussion is beyond the scope of this text.

6.2.6.1 Advantages and disadvantages

The advantages and disadvantages of Cox proportional hazards modeling are very similar to those of logistic regression, discussed in Section 6.2.2.

6.2.7 Neural networks for censored survival data

The theory of ANNs have been discussed in Section 6.2.3. Application to survival requires modification of the ANN output layer to account for right-censored time-event data.

Chi et al. use the ANN architecture proposed by Street [215-216]. The ANN uses the standard three-layered structure. However, in the modified ANN, the survival period is divided into a series of equal intervals from 1 to k, where k is determined by the network designer.

For the purpose of training the ANN, the cumulative probability of survival is calculated for each of the k intervals, for each individual case. The calculated cumulative probabilities are used to assign values to k output neurons. For cases in which an event occurs, the probability of survival is designated 1 in each time interval prior to the event and 0 for all subsequent. For example, if the time period studied is 2 years divided into 4 month intervals, a case in which an event occurs at 17 months would have output neuron values {1, 1, 1, 1, 1, 0, 0, 0}.

Equation 6-27

$$S(t) = \begin{cases} 1, & 0 \le t \le t_i \\ 0, & t > t_i \end{cases}$$

The situation is more complex when censored cases are involved. For cases in which the survival status is known to be alive for the entire study period, the survival probability is designated 1 for each time interval. For cases that are lost to follow-up at time, t, the probability is designated 1 for time intervals prior to time t. After time t, the status of the patient is unknown. The probability of survival is estimated using Kaplan-Meier analysis for each remaining interval. Kaplan-Meier analysis have been discussed in Section 6.2.5. Thus, for censored cases:

Equation 6-28

$$S(t) = \begin{cases} 1, & 0 \leq t \leq t_i \\ \prod \left(\frac{n_j - d_j}{n_j} \right), & t > t_i \end{cases}$$

The resulting trained ANN can be used to estimate the survival probability of novel cases with known explanatory variables. The product of this network structure, diagrammed in Figure 6.6, can be thought of as estimating the survival curve for each case [216].

6.2.8 Recursive partitioning for censored survival data

The basic theory of recursive partitioning implementation for survival analysis does not differ from that which has been previously discussed in Section 6.2.4. The primary modification for censored survival data involves the splitting rule. The Gini coefficient is no longer valid. Instead, a derivative of the Poisson splitting method is used [217]. Initially, the model is represented by:

Equation 6-29

$$\lambda = f(x)$$

where $\boldsymbol{\lambda}$ is an event rate and $f(x)$ represents a list of explanatory variables. Using Poisson model, the splitting rule is based on the deviance, D, for the parent and child Poisson groups:

Equation 6-30

$$Splitting\ Criterion = D_{Parent} - \left(D_{Left\ Child} - D_{Right\ Child} \right)$$

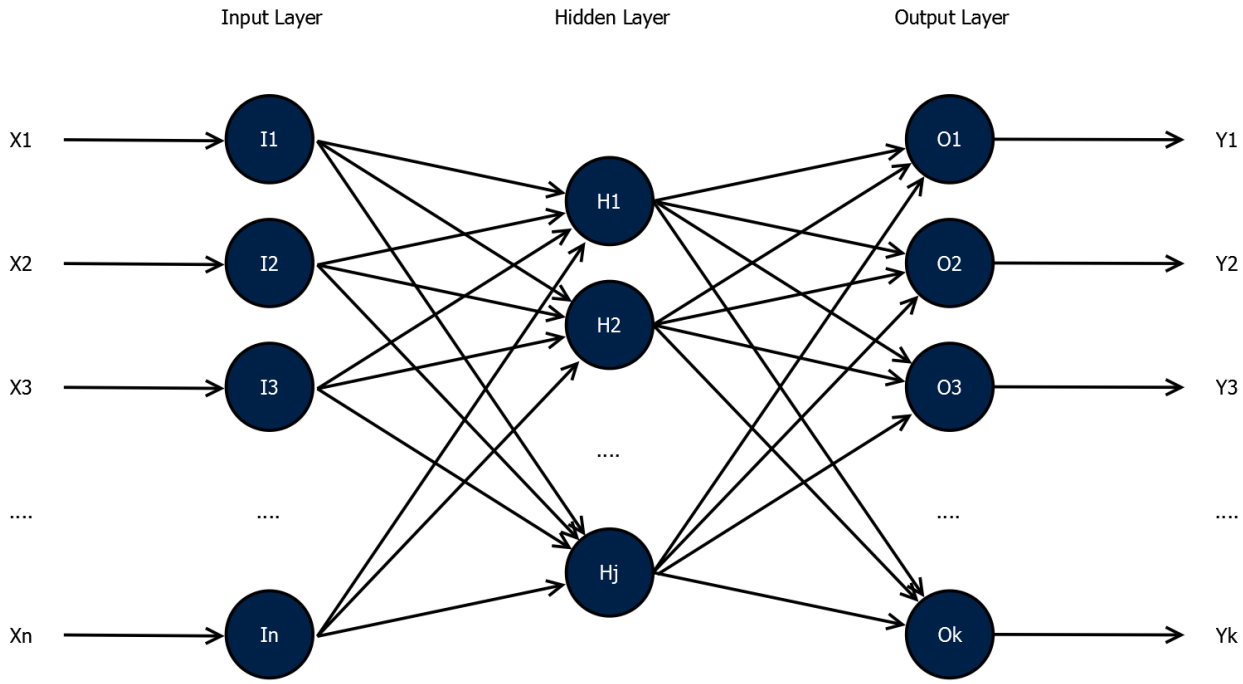


Figure 6.6: Feed-forward artificial neural network for survival. In this modified ANN structure, each output corresponds to the probability of survival at each of k consecutive time intervals. Survival at each time interval is determined by life table analysis.

The deviance within each node is defined as:

Equation 6-31

$$Deviance = \sum \left[c_i \log\left(\frac{c_i}{\lambda t_i}\right) - (c_i - \lambda t_i) \right]$$

where

Equation 6-32

$$\lambda = \frac{\#\ Events}{Total\ Time} = \frac{c_i}{t_i}$$

The Poisson model deals with discrete time data only and must be further modified to handle right-censored survival data [217]. In this modification, the time values are pre-scaled using the exponential scaling of LeBlanc and Crowley [218] .

The survival probability associated with each terminal node is estimated using actuarial life tables for the training cases that comprise each node. When novel data is classified using the tree's splitting rules, the estimated survival probabilities are used to assign survival probabilities to the new cases.

6.2.9 Generalization

Multivariate predictive models must be carefully designed if they are to generalize to external datasets. Poor generalization is often due to over-fitting. When model complexity is very high, or learning is not limited in some way, the modeling algorithm can learn very specific situations in the dataset. When this occurs, the model is sometimes said to be "memorizing" the data. In this situation, the training error of the model is very low, however application to novel data results in high classification error.

Over-fitting is not as pronounced with regression techniques. However, when large numbers of explanatory variables are combined with small sample sizes, over-fitting can result [219].

ANNs are notorious for over-fitting [220]. Anything that increases the complexity of the interconnections of the ANN allows it to model increasingly complex relationships. ANN complexity is enhanced by increasing the number of inputs or the number of neurons in the hidden layer.

Recursive partitioning is also prone to over-fitting [210]. Over-fitting generally increases with tree complexity. For example, a tree could be constructed in which there is one terminal node for each case in the training dataset. This implies 100% accuracy, but generalization to new cases would be poor. Over-fitting is rarely this severe but tree growth must be constrained in some way.

6.2.9.1 Evaluation

To evaluate the generalization, it is necessary to measure the ability of the model to predict the outcome of novel cases. One way to do this is to split the available data into "training" and "testing" subsets. For example, a sample of 1000 cases might be split into 600 training and 400 testing cases. The drawback of this method is that it assumes that are many cases from which to compose the training and testing sets. This is often not the case and, when the sample size is small, splitting can lead to poor statistical power. Even with large samples, the split sample approach is inefficient [219].

An alternative to the split-sample approach is cross-validation. In k-fold cross-validation, the sample is split into k subgroups. For each analysis, one of the k subsamples is used as the testing set while the remaining k-1 subsamples are used to compose the training set. The process is carried out k times until each of the k subsamples has been used exactly once as the testing set. The testing error for each fold is then averaged to give a global measure of classification error and generalization [219].

An extreme, computationally expensive, form of k-fold cross-validation is leave-n-out cross-validation. In this method, n folds are used where n is the sample size [219]. Thus, each testing set consists of exactly one case. Another form of k-fold cross-validation is done by choosing, from the overall dataset, a pre-determined number of random cases, with replacement, as the testing set. The remaining cases are used for the training set. This process can be repeated as many times as desired

and the resulting error measure averaged. The disadvantage of this method is that certain cases may never enter the test data set while others might be selected multiple times [219].

Other methods of assessing the ability of a model to generalize include bootstrap and jack-knife validation. These methods are not discussed here but the reader is referred to Harrell et al. [219].

6.2.9.2 Error measurement

In order to compare model performance, a method for assessing error is required. For classification problems the measure used is usually the classification error:

Equation 6-33

$$Classification\ Error = 1 - Accuracy$$

Alternative methods can also be used, such as sensitivity, specificity, NPV and PPV. The situation is more complex for survival data. One measure is the Brier Score. The Brier Score was originally used by weather forecasters to measure forecasting success by comparing the predicted probability of a weather event to the actual outcome [221].

Equation 6-34

$$Brier\ Score = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2$$

where N refers to the number of "forecasts", f_t refers to the forecasted probability and o_i refers to the actual probability of the outcome. This can be translated to point estimates of survival probabilities where f_t is represented by the model-predicted probability of an event at time *t.* o_t is represented by the actual probability of the outcome at time *t*. For non-censored cases o_t is 1 for those individuals in which an event has occurred prior to time *t*, 0 for those patients known to be alive at time *t*, and a probability calculated by life-table analysis for cases censored prior to time *t*.

6.2.9.3 Prevention

There are several techniques for improving the model generalization. One measure includes careful

selection of the explanatory variables. There is no set rule for determining the number of explanatory variables. However, several "rules of thumb" exist that relate the number of variables to sample size.

It is easy to see the difficulty in designing a model with a hundred predictive factors based a sample of fifty cases. Yet this occurs in many genetic studies. Some authors suggest that the number of explanatory variables should never exceed one tenth of the least-frequent outcome. For example, in a sample of 200 cases of which 150 have outcome A and 50 have outcome B, the recommended number of explanatory variables would be 5. Other authors suggest that the value should be as high as one fifth or as low as one twentieth [219].

The number of explanatory variables can be constrained. Explanatory variables that lack any practical value can be eliminated. Similarly, related variables can be combined or "binned". For example, smoking and drinking can be combined to form single variable [219].

Regression models

Methods of improving the generalization of regression models rely mostly on minimizing the number of explanatory variables and maximizing sample size. Techniques are available that allow explanatory variables to be chosen in such a way that only the most important factors comprise the final model. These techniques include the stepwise inclusion of variables with predictive value or the stepwise elimination of variables that contribute little to prediction [219].

Artificial neural networks

One method of preventing over-fitting in ANNs is called early stopping. In this technique the iterative adjustment of the connection weights is stopped before the training error has reached a minimum. Early stopping can be done manually by specifying the number of iterations before the algorithm stops. Alternatively, the sample can be split into training and validation sets. At each iteration, the error of both the training and validation sets are calculated. Theoretically, the validation dataset error will reach a minimum then begin to increase as over-fitting develops, diagrammed in Figure 6.7. The learning algorithm is stopped when the validation dataset error is at this minimum.

An alternative to early stopping is weight decay. Weight decay introduces a penalty term to the backpropagation algorithm that limits the absolute value of the weight terms. Selecting a weight decay constant is arbitrary. Often, several decay constants are trialed and the value that maximizes generalization is chosen.

The number of neurons in the hidden layer determines model complexity and therefore can contribute to over-fitting. Some "rules of thumb" exist to choose the optimal number of hidden neurons, but most authors tend to choose the number empirically by testing a series of networks, each varying only in the number of hidden neurons.

Recursive partitioning models

Methods for improving the ability of recursive partitioning models to generalize all depend on optimizing tree complexity. Tree growth can be limited by introducing a term to limit both the minimum size of a parent node to be split and the minimum size of a child node to be created. This prevents terminal nodes specific to only a select few cases from being created by the recursive partitioning algorithm. Choice of these parameters is arbitrary and several must be tested by repeated model building [219].

The second commonly used method to limit tree growth is the complexity parameter (CP)., discussed in Section 6.2.4. Like other parameters, the CP is determined empirically [219].

6.3 Methods

In the previous section, a detailed background of the theory of regression, ANNs and recursive partitioning was presented. This section outlines a detailed methodology by which this theory can be applied to oral cancer data. The section begins with variable selection and general procedures. Next, modeling of OCLNM data is discussed, followed by the approach to modeling of survival data.

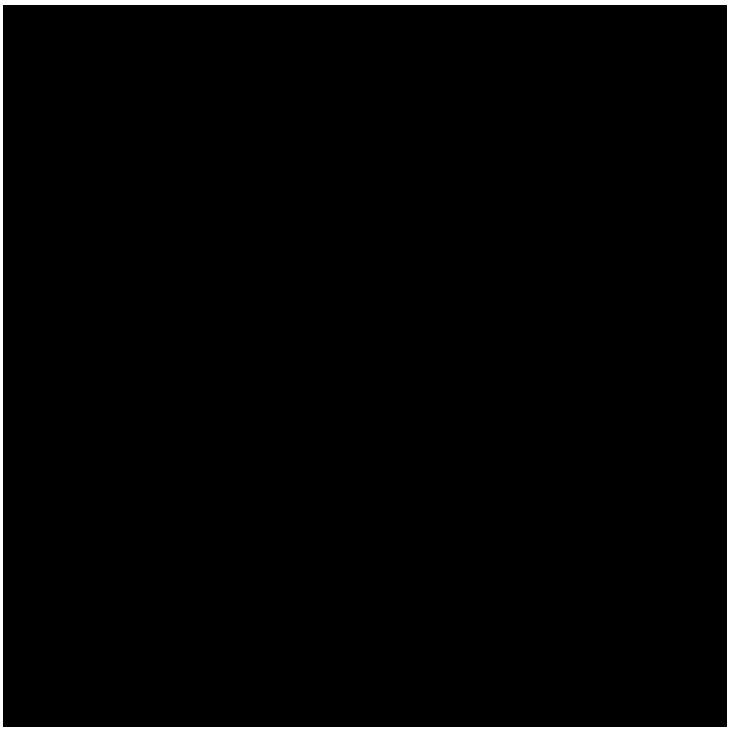


Figure 6.7: Over-fitting in a neural network model. In this example, the error of the training and testing sets both decrease as the model fit improves. However, at iteration 7 over-fitting begins to take place. While the error of the training set continues to decrease, this is due to "memorization" of the data. The over-fitting leads to increased error in the testing dataset. An early stopping method would identify iteration 7 as the point at which to cease further training, optimizing the testing set error.

6.3.1 Data source

All data used in multivariate analysis was from the main database, outlined in Chapter 3, and did not differ from the data sample used in univariate analysis, detailed in Chapter 5. The data is summarized in Appendix IV.

6.3.2 Explanatory variable selection

Explanatory variables were chosen for inclusion in multivariate models based on univariate analysis. Those factors that were significant ($p \leq 0.05$) or approached significance ($0.05 < p \leq 0.10$) were included. Variables were grouped according to availability in the pre-surgical and post-surgical setting. For variables like the T and N categories that have pre-surgical and post-surgical counterparts, only the appropriate version was used (i.e. cT for pre-surgical and pT for port-surgical).

To choose the number of variables used to construct each model, the rule of thumb that the number of explanatory variables should not exceed one tenth of the number of cases in the smallest outcome category was used. For OCLNM, the smallest group was the positive category with n=25. For ACS, DRS and DFS the number of 2-year survival events were 33, 29 and 43 respectively. To ensure a degree of uniformity among models, a maximum number of included explanatory variables was set at 3.

To avoid problems with multicollinearity, models were constrained to contain exactly 1 measurement-based variable. These variables included MRI volume, MRI thickness, MRI diameter, pathological depth, pathological diameter, the log transforms of the aforementioned variables, cT and pT. Similarly, tumour stage was not used as an explanatory variable while the constituent parts, T-stage and N-stage, were included. This arrangement means that the models evaluated all have the explanatory variable structure:

$$\begin{matrix} \boldsymbol{Measurement\ Variable} \\ \boldsymbol{Non-Measurement\ Variable} \\ \boldsymbol{Non-Measurement\ Variable} \end{matrix} \quad \rightarrow \quad \boldsymbol{Outcome}$$

For OCLNM, the variables included in model building are shown in Table 6.2.

Table 6.2: Variables included in OCLNM model building.

Pre-Surgical	*Post-Surgical*
MRI volume + log transform	*MRI volume + log transform*
MRI thickness + log transform	*Pathological depth + log transform*
MRI diameter + log transform	*Pathological diameter + log transform*
cT	*pT*
MRI bone invasion	*Pathological bone invasion*
	LVI
	PNI

For survival analysis, variables with $p \leq 0.10$ in any of 2-year ACS, DRS or DFS were included in model building for all three classes of survival analysis. The variables included are shown in Table 6.3.

Table 6.3: Variables included in survival model building.

Pre-Surgical	*Post-Surgical*
MRI volume + log transform	*MRI volume + log transform*
MRI thickness + log transform	*Pathological depth + log transform*
MRI diameter + log transform	*Pathological diameter + log transform*
cT	*pT*
cN	*pN*
MRI bone invasion	*Pathological bone invasion*
MRI extracapsular spread	*Pathological extracapsular spread*
Age	*Age*
Sex	*Sex*
Midline	*Midline*
ACE-27 Comorbidity	*ACE-27 comorbidity*
	LVI
	PNI
	Grade

6.3.3 General model building procedure

All model building was carried out using R-Project. For each model building technique, an R script was written that first pre-processed the variables appropriately. 10-fold cross-validation was performed.

For OCLNM, predicted probabilities of less than or equal to 0.2 were classified as low-risk for OCLNM while those cases with probabilities greater than 0.2 were classified as moderate-high risk. The model-based outcomes were then compared to the actual outcomes and the NPV calculated. The NPV from each of the 10 folds were averaged and recorded.

For survival, the predicted 2-year survival probabilities for the test set were compared to the survival probability calculated by life table analysis. The Brier Score was then calculated for the test set. The Brier Scores from each of the 10 folds were averaged and recorded.

This process was repeated iteratively for each possible combination of three explanatory variables (adhering to the rule of one measurement-based variable for each model constructed). The general procedure for model building is diagrammed in Figure 6.8.

6.3.4 Logistic regression for occult cervical lymph node metastasis

Logistic regression was carried out using the "glm" (generalized linear model) function included in the R-project software package. For each model constructed, the model intercept and β values for each explanatory variable were extracted and used to calculate the probability of metastasis for each test case using the formula:

Equation 6-35

$$P(metastasis) = \frac{1}{1 + \frac{1}{exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}}$$

6.3.5 Artificial neural network for OCLNM

The ANN was constructed using the "nnet" function of the "Neural Network" R-project software package. A standard three-layer, feed-forward, back-propagated ANN was constructed using weight decay to prevent over-fitting.

The optimum number of neurons in the hidden layer was determined empirically by building each model for specified values from 1 to 5. Similarly, the optimum weight decay value was determined by

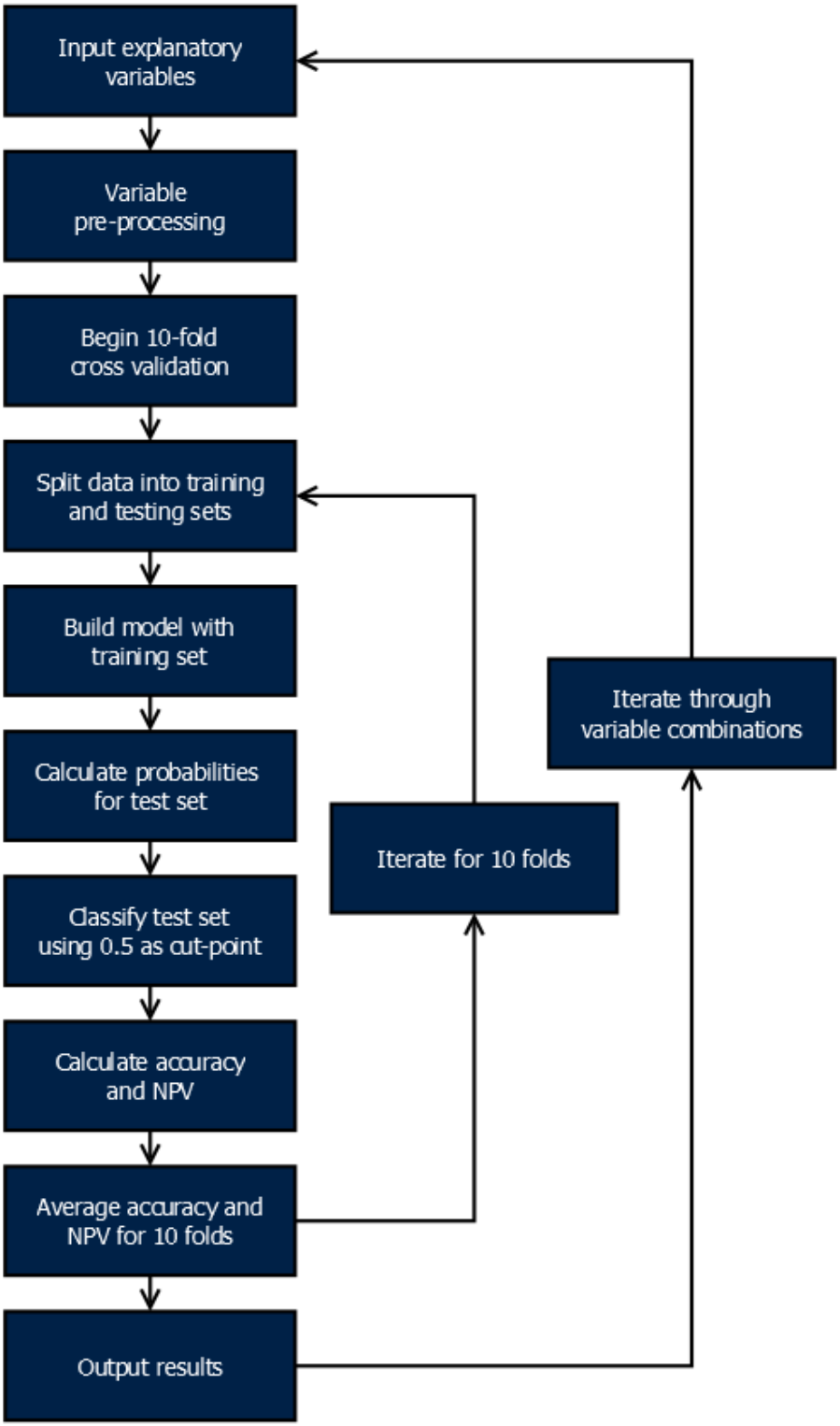


Figure 6.8: General format of model building procedure for OCLNM prediction. An R-project script was written to iterate through variable combinations. 10-fold cross-validation was used to evaluate the ability to generalize.

performing the model building algorithm for each of five possible values 0.001, 0.005, 0.01, 0.05 and 0.1. The number of back-propagation iterations was limited to 100.

The "hidden node" loop and the "weight decay constant" loop were run in such a way that every possible combination of explanatory variable, hidden node value and weight decay constant was evaluated (at great computational expense). The model yielding lowest NPV was taken to be the most proficient. This is diagrammed in Figure 6.9.

6.3.6 Recursive partitioning for occult cervical lymph node metastasis

The recursive partitioning model was constructed using the "rpart" function of the "Recursive Partitioning" R-project software package. The decision tree complexity was limited by specifying 20 as the minimum number of cases in a parent node and 5 cases for the minimum number in a child node. The maximum number of tree levels was 5. The optimum complexity parameter (CP) was determined by iterating through values of 0.001, 0.005, 0.01, 0.05 and 0.1. The model yielding lowest error measure was taken to be the most proficient. This is diagrammed in Figure 6.10.

6.3.7 Comparison of occult cervical lymph node metastasis models

The mean NPV of the testing data sets were used to evaluate each model. The model with the highest NPV was taken as the most proficient model.

6.3.8 Cox proportional hazards for survival

The Cox proportional hazards model was constructed using the "coxph" function of the "Survival" R-Project software package. The probability of survival for cases in the test set were calculated using the "coxph.predict" function, also of the "Survival" package.

6.3.9 Artificial neural network for survival

The basic design of the ANN differed slightly from that used to model OCLNM. The output variable for the training data set was the predicted survival probability at 4, 8, 12, 16, 20 and 24 months.

Like OCLNM, hidden node number and weight decay constant were found by iterating the model

building process through all possible combinations of the parameters. The number of back-propagation iterations was limited to 100.

Survival probabilities at each time interval were calculated using the "cond_prob_nnet" function, a custom R-project script written for this purpose. The "nnet.predict" function of the "Neural Networks" package was used to calculate the predicted survival probability of test cases.

6.3.10 Recursive partitioning for survival

The basic design of the recursive partitioning algorithm for censored survival data did not differ significantly from the structure used for OCLNM prediction.

Like OCLNM, the minimum parent node size was set to 20 and the minimum child node size was set to 5. The optimal CP value was determined by iterating through the possibilities 0.001, 0.005, 0.01, 0.05, 0.1.

Survival probability at each terminal node were determined by using Kaplan-Meier life table analysis on those training cases "binned" in each terminal node. The function "cond_prob" is a custom R-Project script written specifically for this purpose. The 2-year survival probability of the testing set was calculated by using the "rpart.predict" function of the "Recursive Partitioning" package to determine the terminal node to which each test case was assigned. The survival probability at each node was calculated using life table analysis of training cases assigned to each node.

6.3.11 Comparison of survival models

The Brier Score was computed for each model using the Kaplan-Meier survival curve to calculate the probability of survival for each case. The model with the lowest Brier Score was considered the most proficient model.

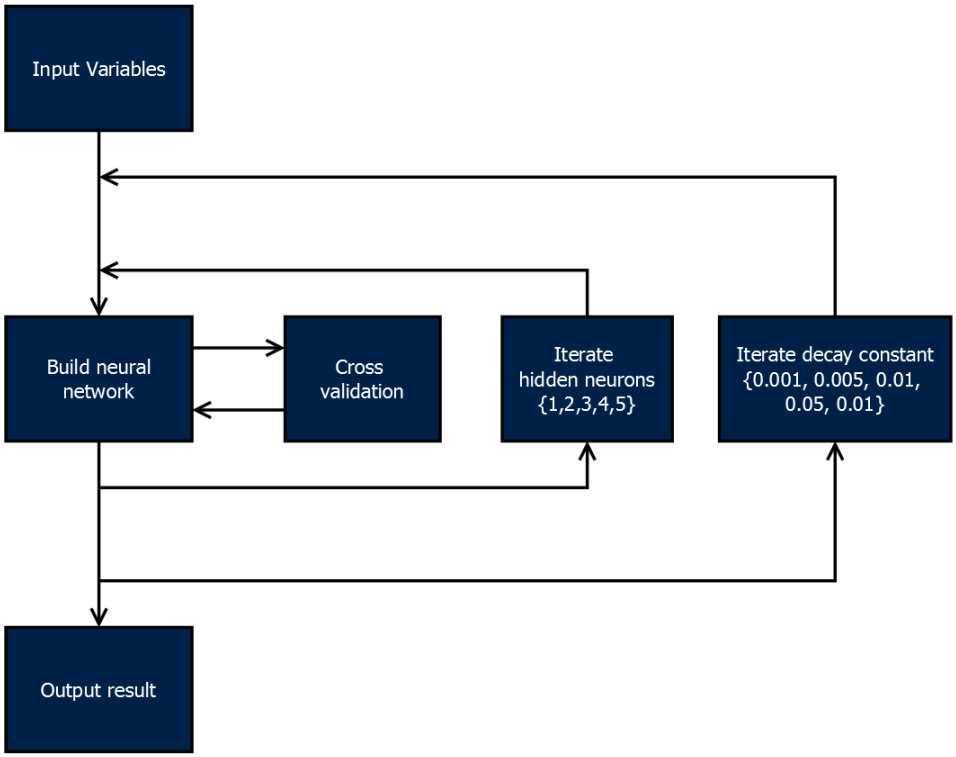


Figure 6.9: Evaluating neural network parameters. The number of hidden neurons and the decay constant must be determined by iterating through all possible combinations with a "nested loop". The process was computationally expensive.

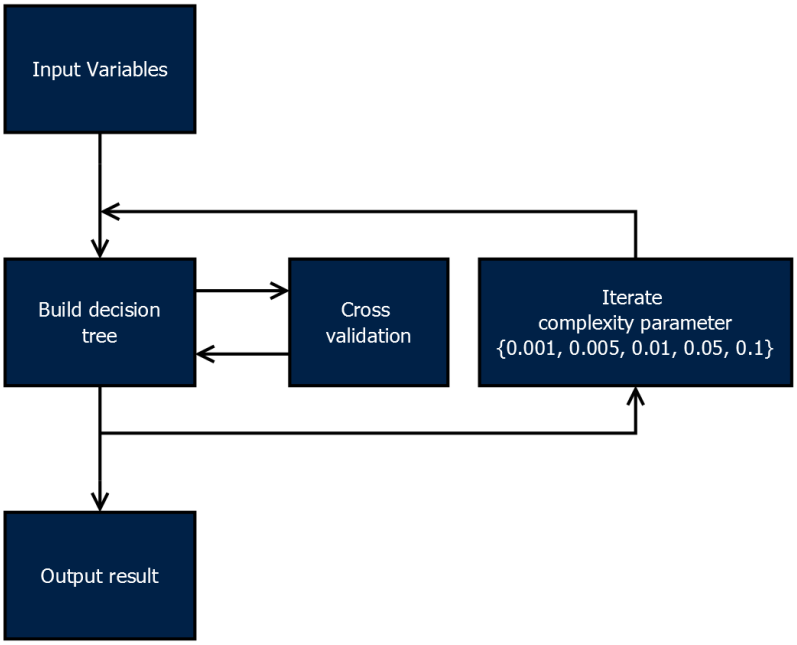


Figure 6.10: Evaluating the recursive partitioning complexity parameter (CP). The CP was determined by iterating through all possibilities.

6.4 Results and discussion: occult cervical lymph node metastasis

This section begins by presenting the results of modeling occult cervical lymph node metastasis. First, models using pre-surgical factors are considered. The results are presented in the following format.

First, a table showing the NPV and composition of the top five models for each of logistic regression, artificial neural networks and recursive partitioning are given. This is followed by the structure and ability to predict OCLNM for each number one ranked model. Structures are presented graphically, with the exception of logistic regression, which is represented by an equation. The ability to classify is represented by a cross-tabulation with associated accuracy, sensitivity, specificity, negative predictive value and positive predictive value.

The results of pre-surgical results are discussed before moving on to the results and discussion of post-surgical modeling. The results and discussion follow the same format as pre-surgical modeling.

For pre-surgical modeling, the results are outlined in Table 6.4. Model structures are found in Equation 6-36, Figure 6.11 and Figure 6.12. The predictive abilities of each model are outlined in Table 6.5, Table 6.6 and Table 6.7.

For post-surgical modeling, the results are outlined in Table 6.8. Model structures are found in Equation 6-37, Figure 6.13 and Figure 6.14. The predictive abilities of each model are outlined in Table 6.9, Table 6.10 and Table 6.11.

6.4.1 Results: pre-surgical prediction

Table 6.4: Models with the 5 highest NPV scores for pre-surgical OCLNM prediction following 10-fold cross-validation for each of Cox proportional hazards, neural network and recursive partitioning methods. The mean, median and range of NPV values are also given for each method. NPV: negative predictive value, CP: complexity parameter, Decay-Nodes: decay constant followed by the number of nodes in the hidden layer.

Logistic Regression (Mean 0.8597, Median 0.8765, Range 0.7183 - 0.9522)

Rank	NPV	Variable 1	Variable 2
1	0.9522	MRI Volume (3 cm^3)	MRI Bone Invasion
1	0.9522	MRI Thickness (11 mm)	MRI Bone Invasion
3	0.9381	MRI Diameter (18.5 mm)	MRI Bone Invasion
4	0.9356	MRI Diameter (log)	MRI Bone Invasion
5	0.8957	MRI Thickness (log)	MRI Bone Invasion

Artificial Neural Network (Mean 0.8875, Median 0.8827, Range 0.8140 - 0.9522)

Rank	NPV	Variable 1	Variable 2	Decay-Nodes
1	0.9522	MRI Volume (3 cm^3)	MRI Bone Invasion	[all]-[all]
1	0.9522	MRI Thickness (11 mm)	MRI Bone Invasion	[all]-[all]
3	0.9381	MRI Diameter (18.5 mm)	MRI Bone Invasion	[all]-[all]
4	0.9208	MRI Diameter (log)	MRI Bone Invasion	0.005-1
5	0.9090	MRI Thickness (log)	MRI Bone Invasion	0.01-1

Recursive Partitioning Tree (Mean 0.8717, Median 0.8556, Range 0.7975 - 0.9522)

Rank	NPV	Variable 1	Variable 2	CP
1	0.9522	MRI Volume (3 cm^3)	MRI Bone Invasion	0.001,0.005,0.01,0.05
1	0.9522	MRI Thickness (11 mm)	MRI Bone Invasion	0.001,0.005,0.01,0.05
2	0.9381	MRI Diameter (18.5 mm)	MRI Bone Invasion	0.001,0.005,0.01
4	0.9300	MRI Thickness	MRI Bone Invasion	0.05
4	0.8901	MRI Thickness (log)	MRI Bone Invasion	0.05

6.4.1.1 Logistic regression

Equation 6-36

$$OR = -2.94 + 1.93 \times Volume\ (3.0cm^3) + 0.33 \times MRI\ Bone\ Invasion$$

with p values of 0.004 for $Volume\ (3.0cm^3)$ and 0.532 for $MRI\ Bone\ Invasion$.

Table 6.5: Cross-tabulation of cases classified according to logistic regression model outcome by actual outcome.

		Outcome		
		Positive	*Negative*	
Test	*Positive*	*22*	*54*	*76*
	Negative	*3*	*57*	*60*
		25	*111*	*136*

Sensitivity	0.88
Specificity	0.51
Negative Predictive Value	0.95
Positive Predictive Value	0.33
Accuracy	0.58

6.4.1.2 Artificial neural networks

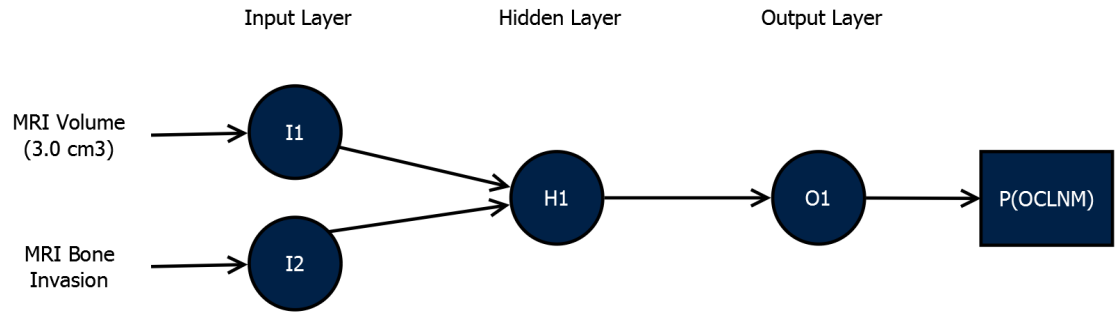


Figure 6.11: Structure of the top-ranked ANN for pre-surgical prediction of OCLNM. The cross-validated NPV result was the same regardless of the number of hidden nodes or the decay constant. The simplest structure is shown.

Table 6.6: Cross-tabulation of cases classified according to artificial neural network outcome by actual outcome.

		Outcome		
		Positive	*Negative*	
Test	*Positive*	*22*	*54*	*76*
	Negative	*3*	*57*	*60*
		25	*111*	*136*

Sensitivity	0.88
Specificity	0.51
Negative Predictive Value	0.95
Positive Predictive Value	0.33
Accuracy	0.58

6.4.1.3 Recursive partitioning

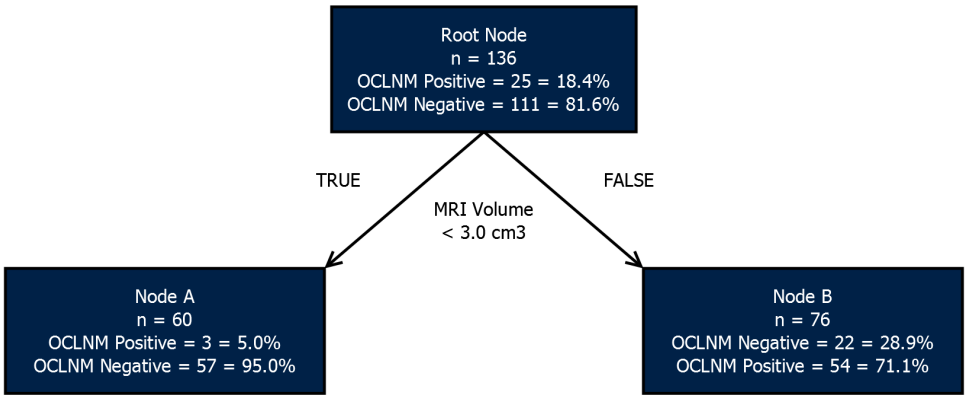


Figure 6.12: Structure of the top-ranked decision tree for pre-surgical prediction of OCLNM. Tumour volume categorized using a single cut-point at 3.0 cm^3 was found to be the most effective criterion. The cross-validated NPV result was the same regardless of CP value.

Table 6.7: Cross-tabulation of cases classified according to recursive partitioning model outcome by actual outcome.

		Outcome		
		Positive	*Negative*	
Test	*Positive*	*22*	*54*	*76*
	Negative	*3*	*57*	*60*
		25	*111*	*136*

Sensitivity	0.88
Specificity	0.51
Negative Predictive Value	0.95
Positive Predictive Value	0.33
Accuracy	0.58

6.4.2 Discussion: pre-surgical prediction

Pre-surgically, MRI-measured volume (3.0 cm^3) and MRI-based thickness (11.0 mm) tied for the number one ranked model in all three modeling methods. This means that volume and thickness, categorized using the median value, are indistinguishable as pre-surgical predictors of OCLNM. Diameter (18.5 mm) was slightly worse, ranking third in each model with an NPV of 0.938. It is tempting to conclude that thickness and volume are more valuable than diameter, but in reality the NPV values were similar.

Imaging-based bone invasion was the only non-measurement variable identified as significant on univariate analysis, limiting the model complexity for pre-surgical OCLNM prediction. As such, each of the 13 models constructed contained imaging-based bone invasion. However, bone invasion was of questionable value. In the logistic regression model, bone invasion had a p-value of 0.532 indicating it was not a significant contributor to the model. In the recursive partitioning model the single parent node was split using volume (3.0 cm^3) without contribution from bone invasion.

Interestingly, the cross-validated NPV for each of logistic regression, ANNs and recursive partitioning analysis was the same at 0.95. Moreover, the ANNs involving volume (3.0 cm^3) and thickness (11.0 mm) had the same NPV regardless of the number of hidden nodes or decay constant. Similarly, the top-ranked recursive partitioning tree was the same regardless of complexity parameter. This likely reflects the overwhelming contribution of the measurement variable to the prediction of pre-surgical OCLNM combined with the questionable contribution of bone invasion. Models need not be structurally complex to predict binary outcomes. There is little opportunity for either of the modeling methods to distinguish themselves based on superior handling of complex variable relationships.

Though 136 cases is a large sample size in OCSCC, the sample size remains small, especially given there were only 25 cases with the event. It is difficult to claim that MRI-based volume is superior to thickness and diameter.

6.4.3 Results: post-surgical prediction

Table 6.8: Models with the 5 highest post-surgical NPV scores for OCLNM prediction following 10-fold cross-validation for each of Cox proportional hazards, neural network and recursive partitioning methods. The mean, median and range of NPV values are also given for each method. LVI: perivascular invasion, PNI: perineural invasion, NPV: negative predictive value, CP: complexity parameter, Decay-Nodes: decay constant followed by the number of nodes in the hidden layer.

Logistic Regression (Mean 0.7748, Median 0.7653, Range 0.7166 - 0.8763)

Rank	NPV	Variable 1	Variable 2	Variable 3
1	0.8763	MRI Volume (0.5, 8 cm^3)	PVI	Path Bone Invasion
2	0.8663	MRI Volume (0.5, 8 cm^3)	PVI	PNI
3	0.8528	MRI Volume (3 cm^3)	PVI	Path Bone Invasion
4	0.8456	MRI Volume (3 cm^3)	PNI	Path Bone Invasion
5	0.8385	MRI Volume (3 cm^3)	PNI	Path Bone Invasion

Artificial Neural Network (Mean 0.8657, Median 0.8631, Range 0.7473 - 0.9439)

Rank	NPV	Variable 1	Variable 2	Variable 3	Decay-Nodes
1	0.9439	MRI Volume (3 cm^3)	PVI	Path Bone Invasion	0.1-[all]
1	0.9425	MRI Volume (3 cm^3)	Tumour Grade	Path Bone Invasion	0.1-[all]
2	0.9414	MRI Volume (3 cm^3)	Tumour Grade	PNI	0.1-[all]
3	0.9353	MRI Volume (3 cm^3)	Tumour Grade	PVI	0.005-3
4	0.9330	MRI Volume (3 cm^3)	PNI	Path Bone Invasion	0.005,0.01-[all]

Recursive Partitioning Tree (Mean 0.8520, Median 0.8519, Range 0.7300 - 0.9575)

Rank	NPV	Variable 1	Variable 2	Variable 3	CP
1	0.9575	MRI Volume (3 cm^3)	PNI	Path Bone Invasion	0.050
1	0.9575	MRI Volume (3 cm^3)	Tumour Grade	Path Bone Invasion	0.050
3	0.9439	MRI Volume (3 cm^3)	Tumour Grade	PNI	0.050
3	0.9439	MRI Volume (0.5, 8 cm^3)	PVI	Path Bone Invasion	0.001,0.005,0.01
5	0.9425	MRI Volume (3 cm^3)	PVI	Path Bone Invasion	0.005,0.01

6.4.3.1 Logistic regression

Equation 6-37

$$OR = -2.80 + 0.62 \times Volume\ (0.5cm^3) + 1.79 \times Volume\ (8.0cm^3) + 1.58 \times PVI + 0.14 \times Pathology\ Bone\ Invasion$$

with p values of 0.459 for $Volume\ (0.5cm^3)$, 0.034 for $Volume\ (8.0cm^3)$, 0.005 for PVI and 0.818 for $Pathology\ Bone\ Invasion$.

Table 6.9: Cross-tabulation of cases classified according to logistic regression model outcome by actual outcome.

		Outcome		
		Positive	*Negative*	
Test	*Positive*	*19*	*32*	*51*
	Negative	*6*	*77*	*83*
		25	*109*	*134*

Sensitivity	0.76
Specificity	0.71
Negative Predictive Value	0.93
Positive Predictive Value	0.37
Accuracy	0.72

6.4.3.2 Artificial neural network

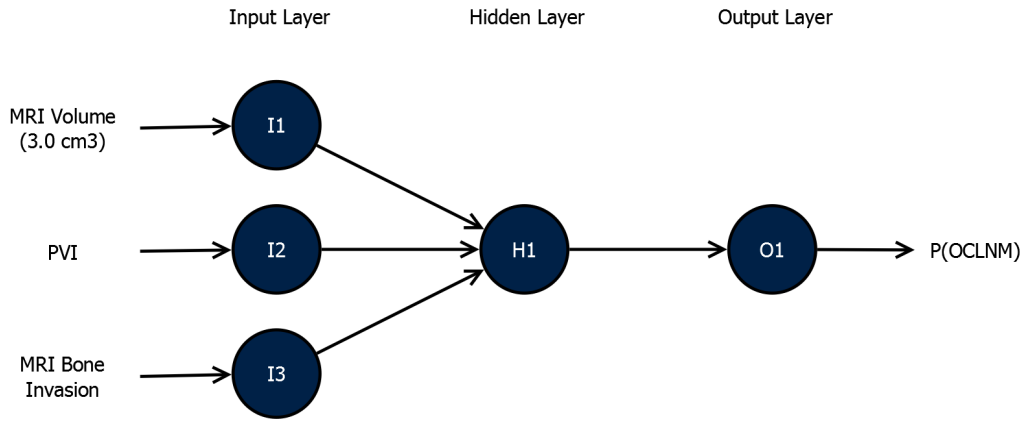


Figure 6.13: Structure of the top-ranked ANN for pre-surgical OCLNM prediction. The cross-validated NPV result was the same regardless of the hidden node number. The decay constant was 0.1. The simplest structure is shown.

Table 6.10: Cross-tabulation of cases classified according to neural network model outcome by actual outcome.

		Outcome		
		Positive	*Negative*	
Test	*Positive*	*22*	*56*	*78*
	Negative	*3*	*53*	*56*
		25	*109*	*134*

Sensitivity	0.88
Specificity	0.49
Negative Predictive Value	0.95
Positive Predictive Value	0.28
Accuracy	0.56

6.4.3.3 Recursive partitioning

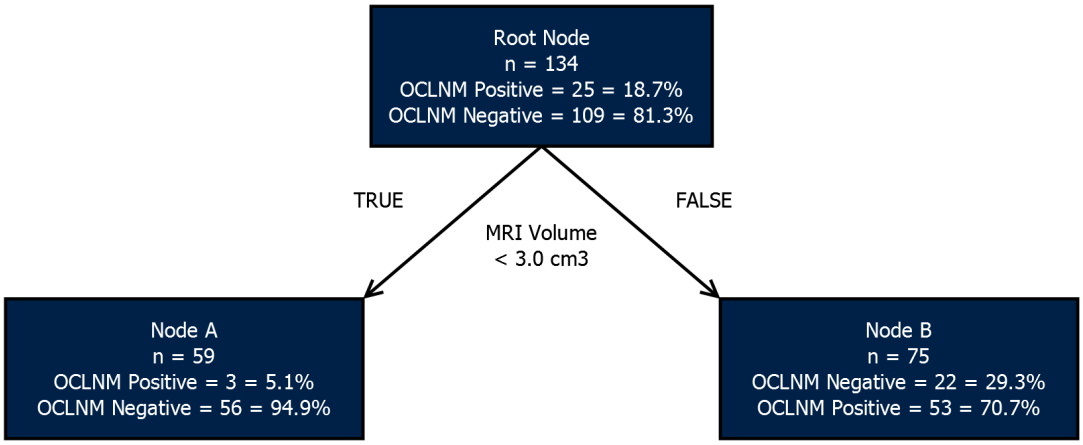


Figure 6.14: Structure of the top-ranked decision tree for pre-surgical prediction of OCLNM. Tumour volume categorized using a single cut-point at 3.0 cm^3 was found to be the most effective criterion. The optimum CP value was 0.05

Table 6.11: Cross-tabulation of cases classified according to neural network model outcome by actual outcome.

		Outcome		
		Positive	*Negative*	
Test	*Positive*	*22*	*53*	*75*
	Negative	*3*	*56*	*59*
		25	*109*	*134*

Sensitivity	0.88
Specificity	0.51
Negative Predictive Value	0.95
Positive Predictive Value	0.29
Accuracy	0.58

6.4.4 Discussion: post-surgical prediction

Post-surgically, more explanatory variables were available and the constructed models more complex. Again, MRI-based volume was found in all of the top-ranked logistic regression, ANN and recursive partitioning models. The top-ranked ANN and recursive partitioning models both used MRI volume (3.0 cm^3), while the top-ranked logistic regression model used MRI volume categorized using the first and third quartiles (0.5, 8 cm^3). There is a strong case that MRI volume is the most important measurement-based predictor of OCLNM.

The remaining variables found to comprise the top-ranked models consisted of various combinations of LVI, PNI, pathological bone invasion and tumour grade. Pathological bone invasion was a component of all three of the number one ranked models. However its contribution to the discriminative ability of the final model is again questionable with a p-value of 0.808. Bone invasion was not included in the top-ranked decision tree.

LVI was an explanatory variable in the top-ranked logistic regression and ANN models. In fact, in the logistic regression model LVI had a p-value of 0.005 and an OR of 1.58. The top-ranked recursive partitioning model was constructed with PNI as a an explanatory variable. However, the final decision tree included only MRI volume (3.0 cm^3). The overall model with the highest NPV at 0.9575 was the recursive partitioning tree based only on MRI volume (3.0 cm^3).

It is surprising that logistic regression and ANNs did not suggest cross-validated models with equivalent NPV values, as occurred in pre-surgical OCLNM prediction. It might be that by forcing the inclusion of three explanatory variables, some over-fitting occurred with the logistic regression and ANN models that did not affect the recursive partitioning model. Constructing models based on two explanatory variables, or a single explanatory variable, might be of some value in clarifying this issue.

A second related issue has to do with the choice of the top-ranked recursive partitioning model. If the final model was based only on MRI volume (3.0 cm^3), why were there only two combinations of variables with an NPV of 0.9575? If MRI volume (3.0 cm^3) was the only contributing variable, then all

models trained using MRI volume (3.0 cm^3) should also have an NPV of 0.9575, regardless of the nature of the second and third explanatory variables. One explanation is that other variable combinations introduced extra splits into the decision tree that increased complexity at the cost of the ability to generalize. The result was a reduction in cross-validated NPV.

As with pre-surgical prediction of OCLNM, sample size was relatively small. The inclusion of additional cases would increase statistical power and enhance the ability of the models to discern increasingly complex variable relationships.

6.5 Results and discussion: survival

In the previous section, the results and discussion of OCLNM modeling were presented. This section follows with the results and discussion for all-cause, disease-related and disease-free survival. Again, pre-surgical and post-surgical models are considered.

The format of the survival results is more complex than that of OCLNM. Again, the section is divided into pre-surgical and post-surgical modeling. However, in each section ACS, DRS and DFS must be considered separately. First, a summary table, similar to those found in the previous section, is presented for ACS, DRS and DFS. As with the previous section, the pre-surgical modeling results are discussed before moving on to the post-surgical modeling results and discussion.

6.5.1 Results: pre-surgical model summaries

For pre-surgical modeling, the results are outlined in Table 6.12, Table 6.13 and Table 6.14. Cox proportional hazards models are given in Equation 6-38 to Equation 6-40. The structure of ANNs are diagrammed in Figure 6.15 to Figure 6.17. The structure of recursive partitioning trees are diagrammed in Figure 6.18 to Figure 6.20. Survival models are compared in Figure 6.21.

6.5.1.1 ACS

Table 6.12: Models with the 5 lowest pre-surgical Briar Scores for 2-year all-cause survival prediction following 10-fold cross-validation for each of Cox proportional hazards, neural network and recursive partitioning methods. The mean, median and range of NPV values are also given for each method. ECS: extracapsular spread, BS: Briar Score, CP: complexity parameter, Decay-Nodes: decay constant followed by the number of nodes in the hidden layer.

Cox (Mean 0.1334, Median 0.1330, Range 0.1214 - 0.1517)

Rank	BS	Variable 1	Variable 2	Variable 3
1	0.1214	MRI Volume (log)	Midline Invasion	cN
2	0.1219	MRI Volume	Midline Invasion	cN
3	0.1219	MRI Volume (log)	Age	Midline Invasion
4	0.1225	MRI Diameter (log)	Midline Invasion	cN
5	0.1225	MRI Volume (log)	MRI ECS	Midline Invasion

Artificial Neural Network (Mean 0.1380, Median 0.1353, Range 0.1205 - 0.1969)

Rank	BS	Variable 1	Variable 2	Variable 3	Decay-Nodes
1	0.1205	MRI Volume	Midline Invasion	cN	0.05-3
2	0.1214	MRI Diam (10,28.5mm)	Midline Invasion	cN	0.001-2
3	0.1215	MRI Volume (log)	Midline Invasion	cN	0.01-3
4	0.1217	MRI Diameter (log)	Midline Invasion	cN	0.005-2
5	0.1222	MRI Thickness (4, 16 mm)	Midline Invasion	cN	0.01-2

Recursive Partitioning Tree (Mean 0.1609, Median 0.1529, Range 0.1188 - 0.2254)

Rank	BS	Variable 1	Variable 2	Variable 3	CP
1	0.1188	MRI Volume (3 cm^3)	Sex	ACE-27	0.01
2	0.1241	MRI Thickness (11 mm)	Sex	ACE-27	0.01
3	0.1309	MRI Volume (3 cm^3)	Age	Sex	0.05
4	0.1314	MRI Volume (3 cm^3)	age	Midline Invasion	0.05
5	0.1316	MRI Volume (3 cm^3)	Sex	Midline Invasion	0.05

6.5.1.2 DRS

Table 6.13: Models with the 5 lowest pre-surgical Briar Scores for 2-year disease-related survival prediction following 10-fold cross-validation for each of Cox proportional hazards, neural network and recursive partitioning methods. The mean, median and range of NPV values are also given for each method. ECS: extracapsular spread, BS: Briar Score, CP: complexity parameter, Decay-Nodes: decay constant followed by the number of nodes in the hidden layer.

Cox (Mean 0.1224, Median 0.1229, Range 0.1072 - 0.1370)

Rank	BS	Variable 1	Variable 2	Variable 3
1	0.1072	MRI Volume	Midline Invasion	cN
2	0.1083	MRI Volume (log)	Midline Invasion	cN
3	0.1088	MRI Thickness	Sex	cN
4	0.1088	MRI Diameter (log)	Midline Invasion	cN
5	0.1092	MRI Thickness	Midline Invasion	cN

Artificial Neural Network (Mean 0.1247, Median 0.1226, Range 0.1021 - 0.1886)

Rank	BS	Variable 1	Variable 2	Variable 3	Decay-Nodes
1	0.1021	MRI Volume (log)	MRI Bone Invasion	cN	0.001-2
2	0.1030	MRI Volume	Midline Invasion	cN	0.01-2
3	0.1068	MRI Thickness (log)	Midline Invasion	cN	0.005-5
4	0.1072	MRI Diam (10, 28.5 mm)	Midline Invasion	cN	0.01-2
5	0.1075	MRI Thickness (log)	Sex	cN	0.005-1

Recursive Partitioning Tree (Mean 0.1424, Median 0.1347, Range 0.1115 - 0.2085)

Rank	BS	Variable 1	Variable 2	Variable 3	CP
1	0.1115	MRI Volume (3 cm^3)	Sex	ACE-27	0.01
2	0.1134	mrthk1cut	Sex	cN	0.01
3	0.1141	cT	Sex	cN	0.01
4	0.1145	mrdiam2cut	Sex	cN	0.05
5	0.1146	mrthk2cut	Sex	cN	0.05

6.5.1.3 DFS

Table 6.14: Models with the 5 lowest pre-surgical Briar Scores for 2-year disease-related survival prediction following 10-fold cross-validation for each of Cox proportional hazards, neural network and recursive partitioning methods. The mean, median and range of NPV values are also given for each method. ECS: extracapsular spread, BS: Briar Score, CP: complexity parameter, Decay-Nodes: decay constant followed by the number of nodes in the hidden layer.

Cox (Mean 0.1660, Median 0.1655, Range 0.1570 - 0.1815)

Rank	BS	Variable 1	Variable 2	Variable 3
1	0.1570	MRI Volume (log)	MRI Bone Invasion	Midline Invasion
2	0.1576	MRI Volume (log)	MRI Bone Invasion	Sex
3	0.1580	MRI Volume (log)	Sex	Midline Invasion
4	0.1584	MRI Diameter	Sex	Midline Invasion
5	0.1584	MRI Volume (log)	MRI Bone Invasion	Age

Artificial Neural Network (Mean 0.1704, Median 0.1673, Range 0.1433 - 0.2359)

Rank	BS	Variable 1	Variable 2	Variable 3	Decay-Nodes
1	0.1433	MRI Diameter (18.5mm)	Midline Invasion	cN	0.05-5
2	0.1449	MRI Volume	MRI Bone Invasion	cN	0.01-5
3	0.1489	MRI Volume (3 cm^3)	Midline Invasion	cN	0.01-2
4	0.1490	MRI Diameter (18.5mm)	MRI Bone Invasion	Midline Invasion	0.1-5
5	0.1491	MRI Diameter (18.5mm)	Sex	cN	0.1-5

Recursive Partitioning Tree (Mean 0.1924, Median 0.1821, Range 0.1539 - 0.2722)

Rank	BS	Variable 1	Variable 2	Variable 3	CP
1	0.1472	MRI Volume	MRI Bone Invasion	Age	0.05
2	0.1500	MRI Volume (log)	MRI Bone Invasion	Age	0.05
3	0.1506	MRI Volume	MRI Bone Invasion	Sex	0.05
4	0.1510	MRI Volume (log)	MRI Bone Invasion	Sex	0.05
5	0.1510	MRI Volume	MRI Bone Invasion	Midline Invasion	0.05

6.5.2 Results: pre-surgical Cox proportional hazards

6.5.2.1 ACS

Equation 6-38

$$HR = 1.67\times MRI\ Volume\ (\log \mathrm{X}) + 0.56\times Midline\ (Abuts) + 3.52\times Midline\ (Crosses) + 3.23\times cN\ (1) + 22.31\times cN\ (2)$$

with p values of 0.009 for $MRI\ Volume\ (\log)$, 0.232 for $Midline\ (Abuts)$, 0.031 for $Midline\ (Crosses)$, 0.005 for $cN\ (1)$ and 0.001 for $cN\ (2)$.

6.5.2.2 DRS

Equation 6-39

$$HR = 1.03\times MRI\ Volume + 0.69\times Midline\ (Abuts) + 4.87\times Midline\ (Crosses) + 6.21\times cN\ (1) + 46.55\times cN\ (2)$$

with p values of 0.02 for $MRI\ Volume$, 0.438 for $Midline\ (Abuts)$, 0.001 for $Midline\ (Crosses)$, <0.0001 for $cN\ (1)$ and <0.00001 for $cN\ (2)$.

6.5.2.3 DFS

Equation 6-40

$$HR = 3.62\times MRI\ Volume\ (\log \mathrm{x}) + 0.85\times Midline\ (Abuts) + 1.91\times Midline\ (Crosses) + 1.05 \times MRI\ Bone\ Invasion$$

with p values of 0.005 for $MRI\ Volume\ (\log \mathrm{x})$, 0.714 for $Midline\ (Abuts)$, 0.271 for $Midline\ (Crosses)$, 0.893 for $MRI\ Bone\ Invasion$.

6.5.3 Results: pre-surgical artificial neural networks

6.5.3.1 ACS

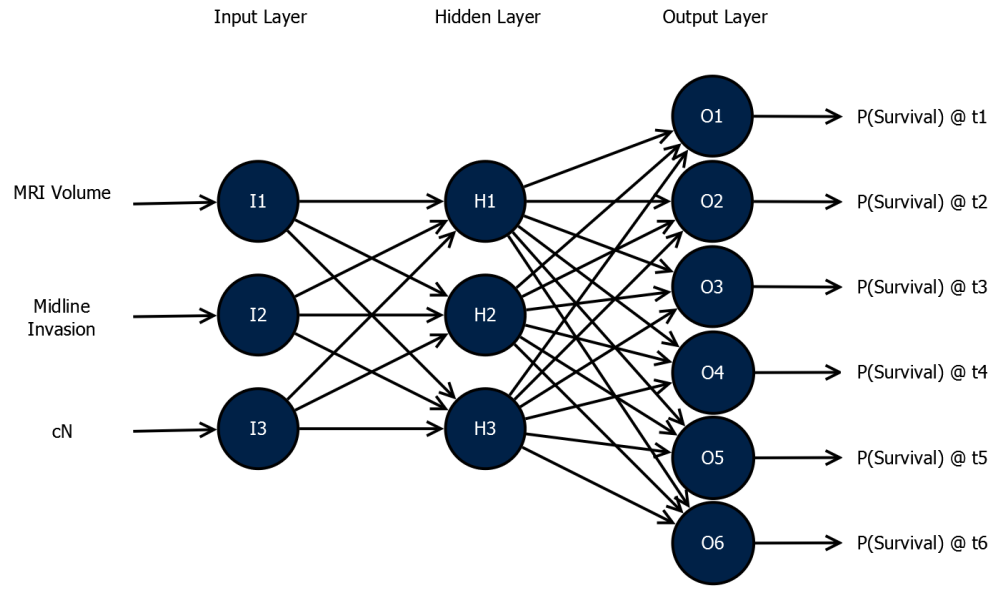


Figure 6.15: Structure of the top-ranked ANN for pre-surgical prediction of 2-year all-cause survival. The explanatory variables were passed to three hidden nodes. The decay constant was 0.05. Each node in the output layer represents the survival probability at six equal time intervals {t1, t2, t3, t4, t5, t6}.

6.5.3.2 DRS

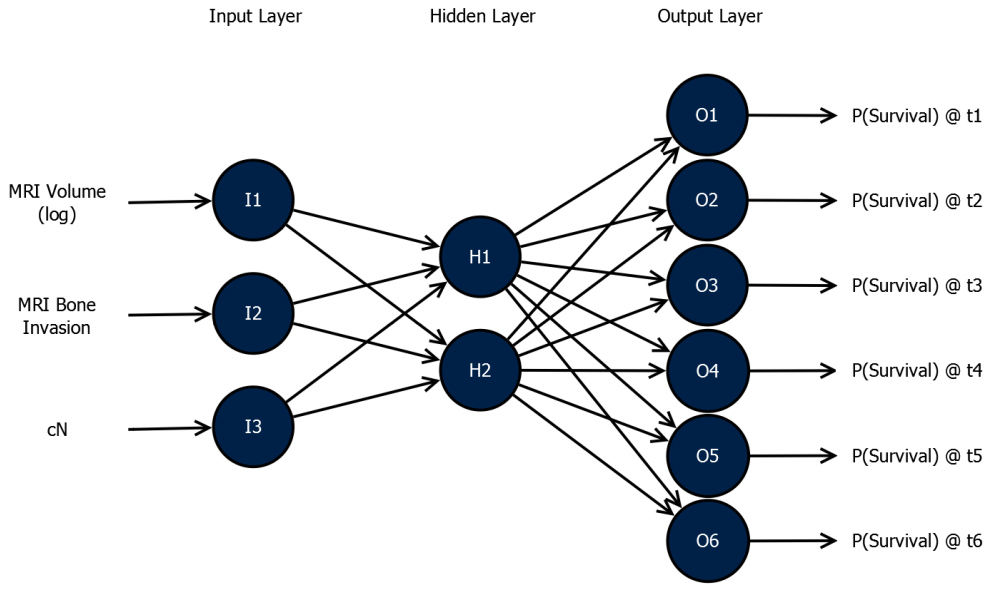


Figure 6.16: Structure of the top-ranked ANN for pre-surgical prediction of 2-year disease-related survival. The explanatory variables were passed to two hidden nodes. The decay constant was 0.001. Each node in the output layer represents the survival probability at six equal time intervals {t1, t2, t3, t4, t5, t6}.

6.5.3.3 DFS

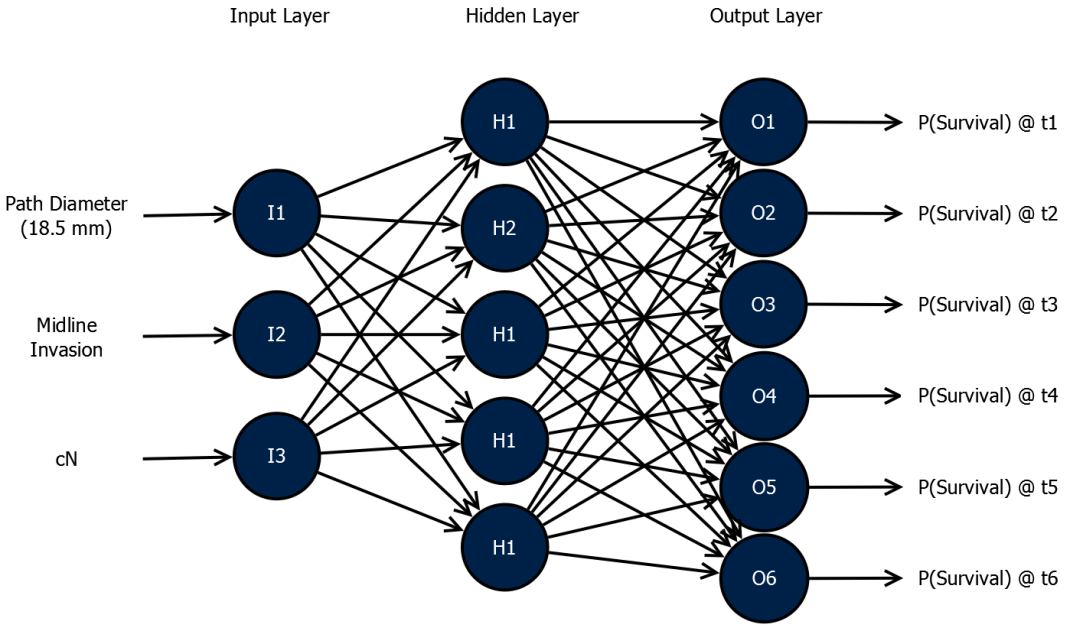


Figure 6.17: Structure of the top-ranked ANN for pre-surgical prediction of 2-year disease-free survival. The explanatory variables were passed to five hidden nodes. The decay constant was 0.05. Each node in the output layer represents the survival probability at six equal time intervals {t1, t2, t3, t4, t5, t6}.

6.5.4 Results: pre-surgical recursive partitioning

6.5.4.1 ACS

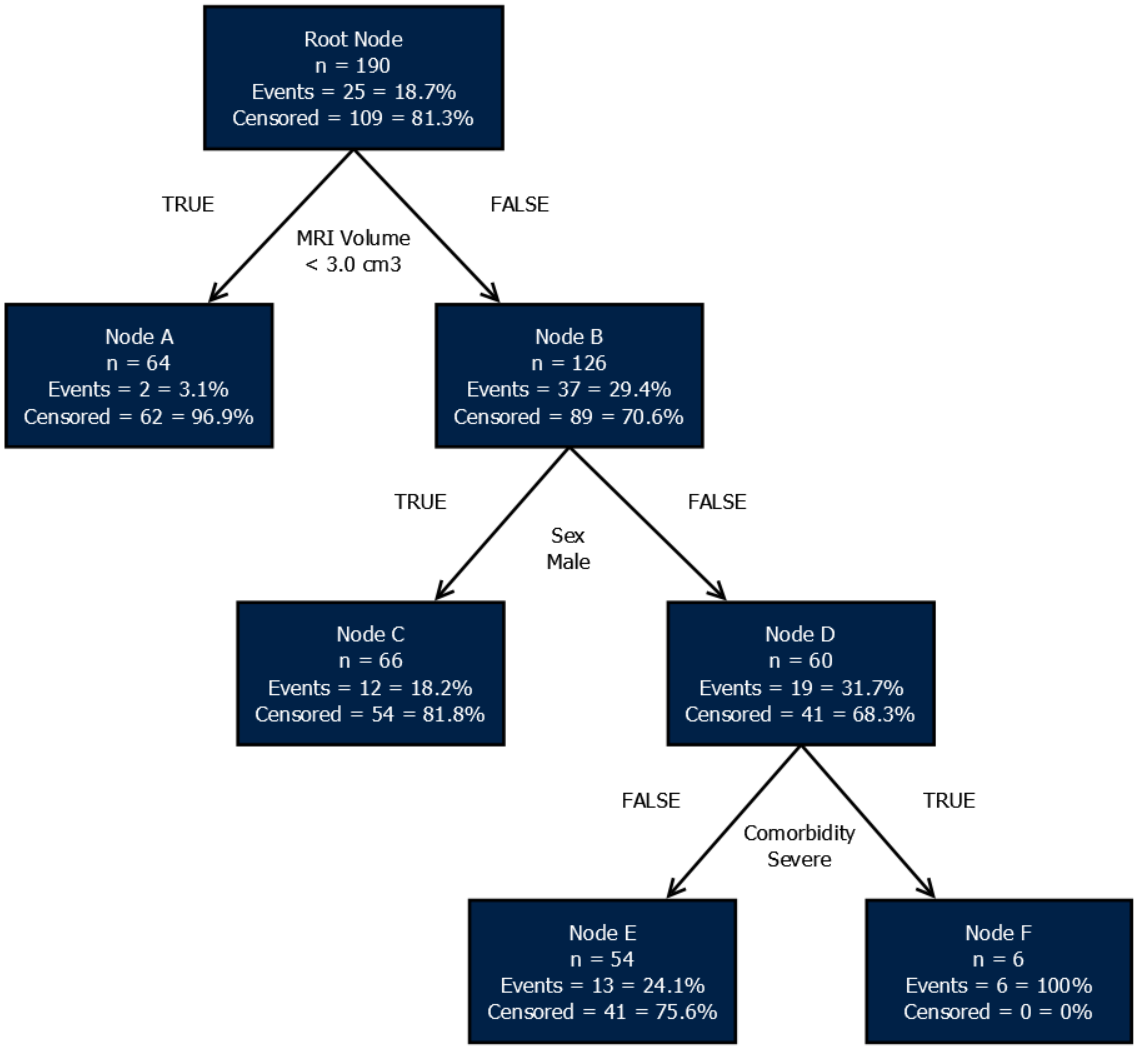


Figure 6.18: Structure of the top-ranked decision tree for pre-surgical prediction of 2-year all-cause survival. The final tree consisted of 3 levels and had 4 terminal nodes. The complexity parameter was 0.01.

6.5.4.2 DRS

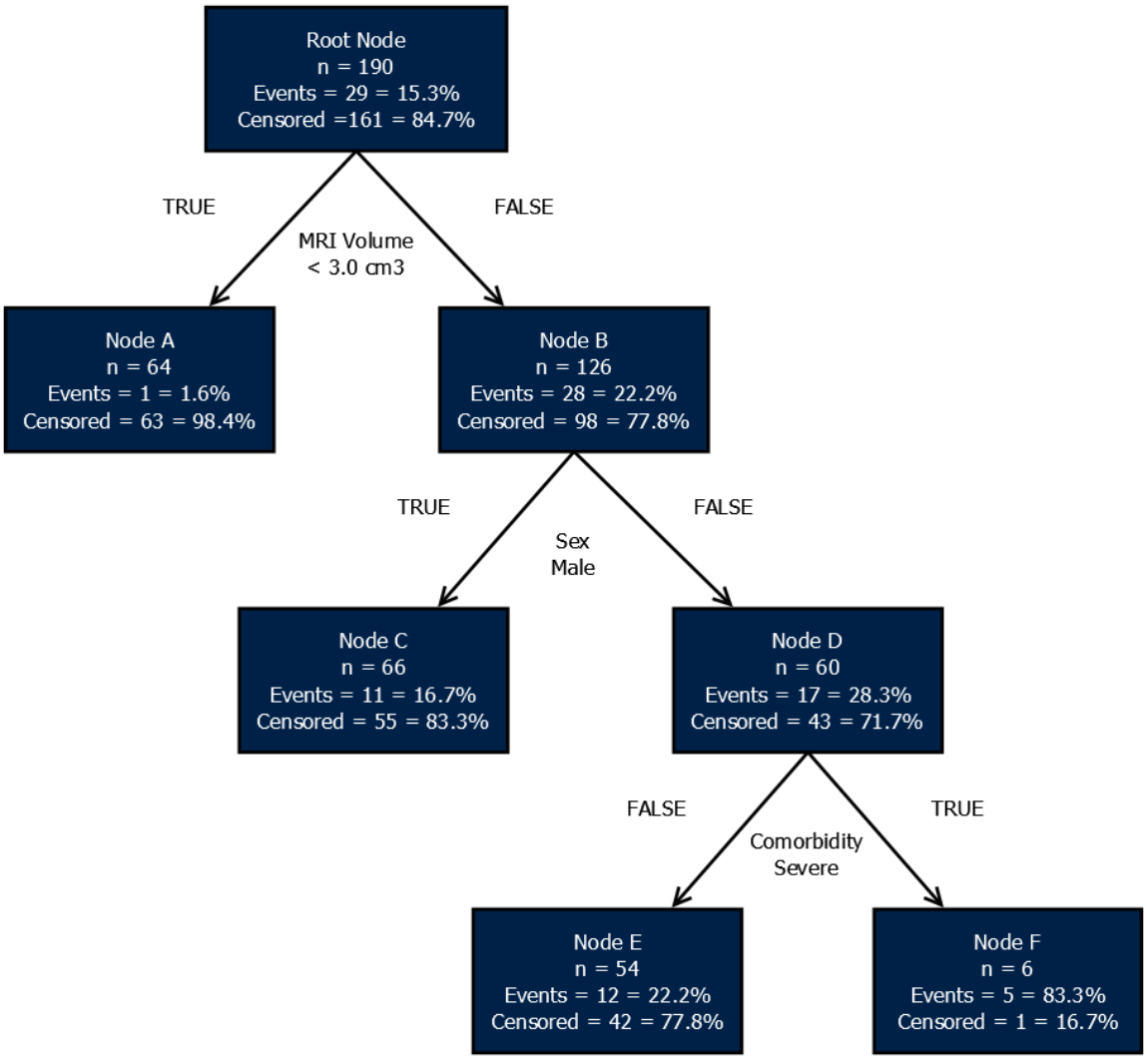


Figure 6.19: Structure of the top-ranked decision tree for pre-surgical prediction of 2-year disease-free survival. The final tree consisted of 3 levels and had 4 terminal nodes. The complexity parameter was 0.01.

6.5.4.3 DFS

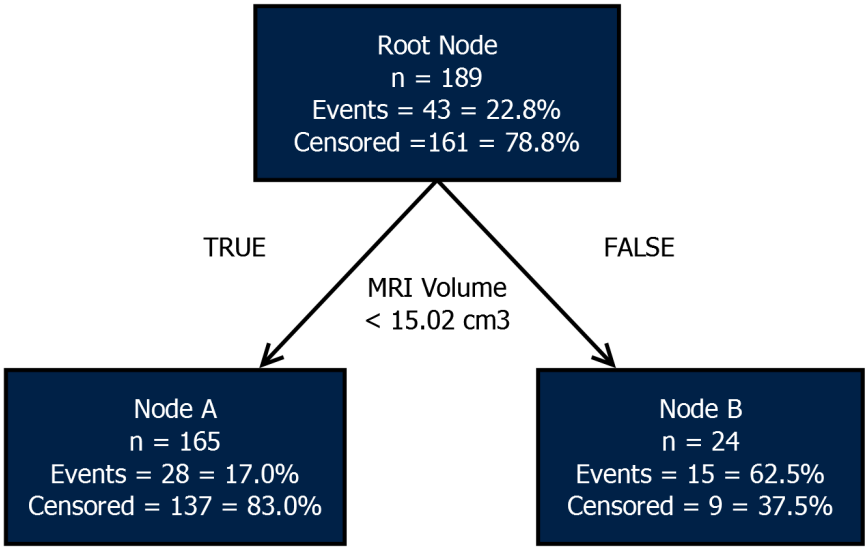


Figure 6.20: Structure of the top-ranked decision tree for pre-surgical prediction of 2-year disease-related survival. The decision tree used a cut-point of 15.0 cm^3 for continuous volume. The final tree consisted of 1 level and had 2 terminal nodes. The complexity parameter was 0.05.

6.5.5 Results: pre-surgical model comparison

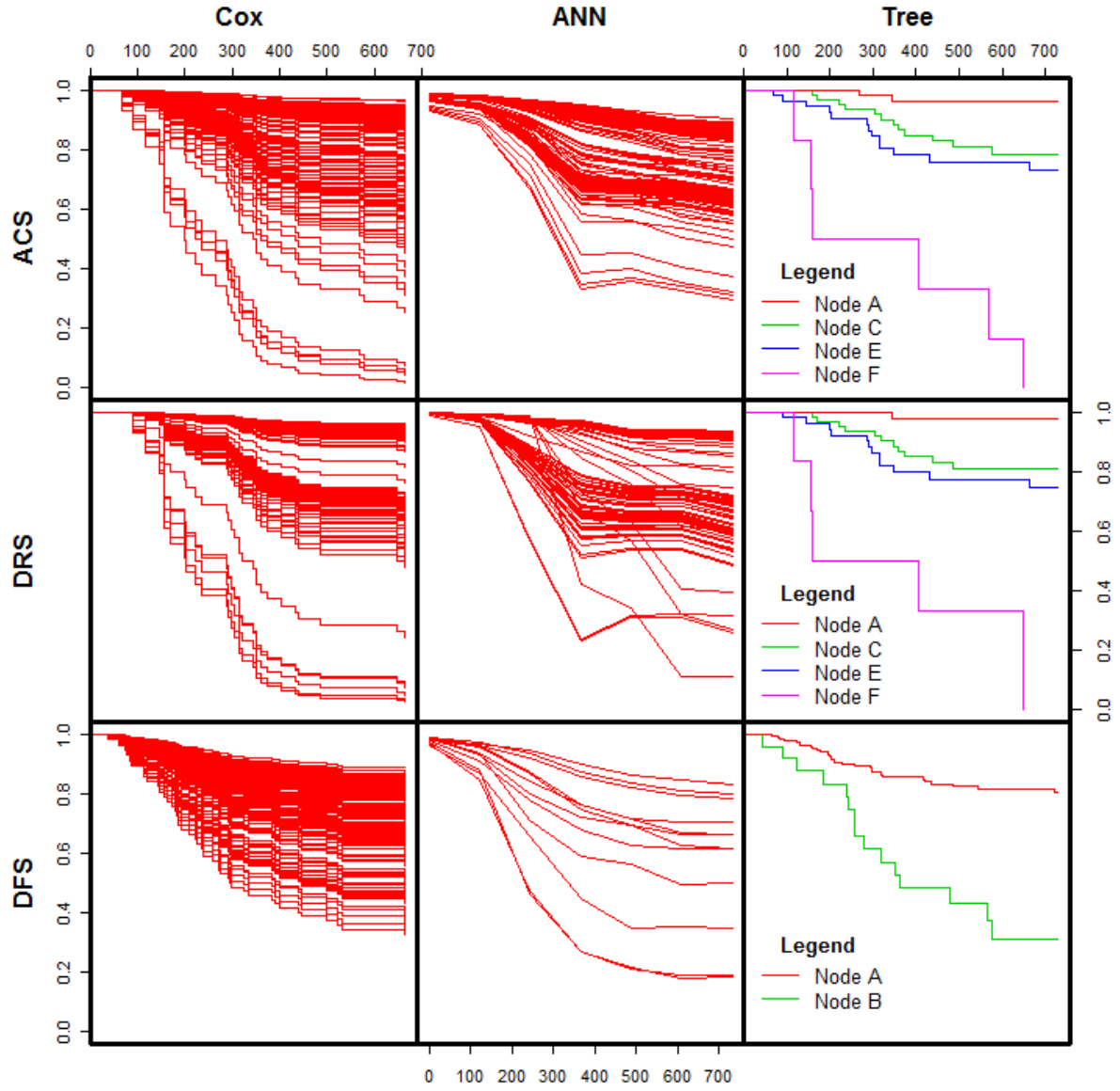


Figure 6.21: Plot of 2-year survival curves for the three pre-surgical modeling methods. The first column is constructed from the pre-surgical Cox model using the "survfit" function of R-Project "Survival" package. The function estimates the baseline hazard of the sample then uses the hazard ratio of each variable combination to plot the estimated survival curves. The middle column is constructed from the pre-surgical artificial neural network output. The model output estimates the survival probability at six equal time intervals. In the plot, each survival curve represents a distinct combination of values of the three explanatory variables. The right column is constructed from the pre-surgical recursive partitioning output. The cases were categorized according to the decision tree and the Kaplan-Meier curve plotted for each category. The vertical axis represents cumulative survival while the horizontal axis represents survival time in days. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival; Cox: Cox proportional hazards; ANN: artificial neural network; tree; recursive partitioning decision tree.

6.5.6 Discussion: pre-surgical prediction

For ACS, the top-ranked variable combination for both the Cox and ANN methods were MRI volume, midline invasion and cN. It was expected that cN would play a large role in any model given its association with a 50% increase in mortality. cN1 and cN2 tumours were highly significant contributors to the Cox model with HRs of 3.23 and 22.31 respectively.

Midline invasion also makes clinical sense as a survival predictor. Tumours crossing the midline are large and can metastasize contralaterally. Tumours crossing the midline significantly contributed to the Cox model with a HR of 3.52, while those abutting the midline had a HR of 0.56. While the result for tumour abutting the midline was significant, the HR implies that these tumours have a survival advantage relative to those that do not approach the midline. This makes little clinical sense and must be interpreted with caution.

MRI-based volume was the measurement variable in the top-ranked model for each of the three ACS modeling methods. This is likely a reflection of the role of volume in CLNM prediction, as outlined in the previous sections. However, the results suggest that volume also has independent value beyond CLNM prediction given that cN was included in the top-ranked models as well.

Volume of invasion showed value beyond ACS. Volume was a component of all three top-ranked DRS models, as well as the top-ranked Cox and recursive partitioning model for DFS, which MRI Diameter (18.5 mm). Volume is clearly superior to MRI-based thickness and diameter for pre-surgical survival prediction.

The remainder of the variables included in top-ranked pre-surgical models included MRI bone invasion and MRI midline invasion. These structures add an anatomic structure-based element to the prognostic models, much like critical structures play a role in T4a classification. cN was a prominent feature of all of the top-ranked Cox and ANN models, with the exception of the top-ranked DFS Cox model . This is not surprising and reflects the association of CLNM with a 50% increase in mortality.

The results of the recursive partitioning analyses did not agree with their Cox and ANN counterparts. For ACS and DRS the top-ranked recursive partitioning model featured MRI volume, sex and ACE-27 comorbidity. The inclusion of volume is not surprising as it has already been shown to predict OCLNM and thus has survival prediction value. However, the inclusion of the ACE-27 in DRS prediction is somewhat surprising. It could be that the presence of comorbidity results in less aggressive treatment, late presentation or affects other factors related to DRS that have not been considered. There is little clinical evidence that sex is a strong predictor of survival in oral cancer.

Because this result makes little clinical sense, it should be regarded with suspicion. Equally suspicious is the absence of cN in the ACS and DRS models. cN is a strong predictor of survival and should be a vital component of any survival model. Its exclusion here may reflect the ability of the combination of MRI volume, sex and ACE-27 comorbidity to act as a "surrogate" for the predictive value of cN.

The results of recursive partitioning for RFS were slightly different with MRI volume, age and imaging bone invasion. However, the structure of the decision tree reveals that only MRI-based volume was a significant predictor of outcome. It is surprising that cN did not contribute to the prediction of DFS and is another reason why recursive partitioning should be regarded with suspicion.

6.5.7 Results: post-surgical model summaries

For post-surgical modeling, the results are outlined in Table 6.15, Table 6.16 and Table 6.17. Cox proportional hazards models are given in Equation 6-41 to Equation 6-43. The structure of ANNs are diagrammed in Figure 6.22 to Figure 6.24. The structure of recursive partitioning trees are diagrammed in Figure 6.25 to Figure 6.27. Survival models are compared in Figure 6.28.

6.5.7.1 ACS

Table 6.15: Models with the 5 lowest post-surgical Briar Scores for 2-year all-cause survival prediction following 10-fold cross-validation for each of Cox proportional hazards, neural network and recursive partitioning methods. The mean, median and range of NPV values are also given for each method. ECS: extracapsular spread, BS: Briar Score, CP: complexity parameter, Decay-Nodes: decay constant followed by the number of nodes in the hidden layer.

Cox (Mean 0.1303, Median 0.1311, Range 0.1127 - 0.1479)

Rank	*BS*	*Variable 1*	*Variable 2*	*Variable 3*
1	0.1113	MRI Volume	Midline Invasion	pN
2	0.1127	Path Depth (6 mm)	Midline Invasion	pN
3	0.1134	Path Diameter (10, 25mm)	Grade	pN
4	0.1142	MRI Volume (0.5, 8 cm^3)	Grade	pN
5	0.1148	Path Depth (6 mm)	PVI	Midline Invasion

Artificial Neural Network (Mean 0.1314, Median 0.1304, Range 0.1052 - 0.1927)

Rank	*BS*	*Variable 1*	*Variable 2*	*Variable 3*	*Decay-Nodes*
1	0.1052	MRI Volume (0.5, 8 cm^3)	PVI	Bone Invasion	0.005-2
2	0.1053	Path Depth (2, 10 mm)	Midline Invasion	pN	0.01-2
3	0.1056	MRI Volume	Midline Invasion	pN	0.005-1
4	0.1082	Path Depth (2, 10 mm)	Age	pN	0.05-3
5	0.1094	Path Depth (2, 10 mm)	PVI	Midline Invasion	0.05-4

Recursive Partitioning Tree (Mean 0.1537, Median 0.1449, Range 0.1188 - 0.2273)

Rank	*BS*	*Variable 1*	*Variable 2*	*Variable 3*	*CP*
1	0.1188	MRI Volume (3 cm^3)	Sex	ACE-27 Comorbidity	0.01
2	0.1194	Path Diameter (10, 25mm)	Midline Invasion	pN	0.001
3	0.1203	MRI Volume (3 cm^3)	Path Bone Invasion	Midline Invasion	0.005
4	0.1206	pT	LVI	Midline Invasion	0.05
5	0.1206	pT	PNI	Midline Invasion	0.05

6.5.7.2 DRS

Table 6.16: Models with the 5 lowest post-surgical Briar Scores for 2-year disease-related survival prediction following 10-fold cross-validation for each of Cox proportional hazards, neural network and recursive partitioning methods. The mean, median and range of NPV values are also given for each method. ECS: extracapsular spread, BS: Briar Score, CP: complexity parameter, Decay-Nodes: decay constant followed by the number of nodes in the hidden layer.

Cox (Mean 0.1168, Median 0.1182, Range 0.0919 - 0.1368)

Rank	BS	Variable 1	Variable 2	Variable 3
1	0.0919	MRI Volume	Midline Invasion	pN
2	0.0938	Path Diameter (10, 25mm)	Grade	pN
3	0.0949	Path Depth (6 mm)	Midline Invasion	pN
4	0.0960	Path Depth (6 mm)	Grade	pN
5	0.0963	MRI Volume (0.5, 8 cm^3)	Grade	pN

Artificial Neural Network (Mean 0.1177, Median 0.1174, Range 0.0857 - 0.1797)

Rank	BS	Variable 1	Variable 2	Variable 3	Decay-Nodes
1	0.0857	MRI Volume	Midline Invasion	pN	0.001-1
2	0.0880	Path Depth (2, 10 mm)	Midline Invasion	pN	0.05-4
3	0.0906	Path Depth (2, 10 mm)	Age	pN	0.05-3
4	0.0907	Path Depth (log)	Midline Invasion	pN	0.001-3
5	0.0908	Path Diameter (10, 25mm)	Midline Invasion	pN	0.005-1

Recursive Partitioning Tree (Mean 0.1537, Median 0.1449, Range 0.1188 - 0.2273)

Rank	BS	Variable 1	Variable 2	Variable 3	CP
1	0.1025	MRI Volume (3 cm^3)	Path ECS	Midline Invasion	0.1
1	0.1025	MRI Volume (3 cm^3)	LVI	Midline Invasion	0.05,0.1
1	0.1025	MRI Volume (3 cm^3)	PNI	Midline Invasion	0.05,0.1
1	0.1025	Path Diameter (10, 25mm)	Path ECS	Midline Invasion	0.1
1	0.1025	Path Diameter (10, 25mm)	LVI	Midline Invasion	0.05,0.1

6.5.7.3 DFS

Table 6.17: Models with the 5 lowest post-surgical Briar Scores for 2-year disease-free survival prediction following 10-fold cross-validation for each of Cox proportional hazards, neural network and recursive partitioning methods. The mean, median and range of NPV values are also given for each method. ECS: extracapsular spread, BS: Briar Score, CP: complexity parameter, Decay-Nodes: decay constant followed by the number of nodes in the hidden layer.

Cox (Mean 0.1615, Median 0.1617, Range 0.1407 - 0.1812)

Rank	BS	Variable 1	Variable 2	Variable 3
1	0.1407	Path Diameter (10, 25mm)	Grade	pN
2	0.1411	Path Depth (6 mm)	Grade	pN
3	0.1420	Path Diameter (15.5mm)	Grade	pN
4	0.1449	MRI Volume (3 cm^3)	Grade	pN
5	0.1453	Path Depth (6 mm)	Midline Invasion	pN

Artificial Neural Network (Mean 0.1597, Median 0.1578, Range 0.1325 - 0.1927)

Rank	BS	Variable 1	Variable 2	Variable 3	Decay-Nodes
1	0.1325	Path Diameter (10, 25mm)	Midline Invasion	pN	0.005-1
2	0.1349	MRI Volume	PVI	Grade	0.01-3
3	0.1360	Path Diameter (10, 25mm)	Grade	pN	0.05-1
4	0.1382	Path Depth (2, 10 mm)	Age	pN	0.001-1
5	0.1384	Path Depth (6 mm)	Grade	pN	0.05-1

Recursive Partitioning Tree (Mean 0.1537, Median 0.1449, Range 0.1188 - 0.2273)

Rank	BS	Variable 1	Variable 2	Variable 3	CP
1	0.1400	Path Diameter (15.5mm)	LVI	Grade	0.005
2	0.1466	Path Depth (3, 10mm)	Sex	LVI	0.05
2	0.1466	Path Depth (3, 10mm)	LVI	PNI	0.05
2	0.1466	Path Depth (3, 10mm)	Sex	Midline Invasion	0.05
2	0.1466	Path Depth (3, 10mm)	LVI	ACE-27 Comorbidity	0.05

6.5.8 Results: post-surgical Cox proportional hazards

6.5.8.1 ACS

Equation 6-41

$$RR = 1.05\times MRI\ Volume + 0.36\times Midline\ (Abuts) + 3.82\times Midline\ (Crosses) + 6.90\times pN\ (1) + 11.44\times pN\ (2)$$

with p values of <0.001 for $MRI\ Volume$, 0.057 for $Midline\ (Abuts)$, 0.016 for $Midline\ (Crosses)$, <0.001 for $cN\ (1)$ and <0.000001 for $cN\ (2)$.

6.5.8.2 DRS

Equation 6-42

$$RR = 1.06\times MRI\ Volume + 0.35\times Midline\ (Abuts) + 3.31\times Midline\ (Crosses) + 13.97\times pN\ (1) + 23.25\times pN\ (2)$$

with p values of <0.001 for $MRI\ Volume$, 0.056 for $Midline\ (Abuts)$, 0.061 for $Midline\ (Crosses)$, <0.001 for $pN\ (1)$ and <0.0000001 for $pN\ (2)$.

6.5.8.3 DFS

Equation 6-43

$$RR = 0.57\times Path\ Diameter\ (10\ \mathrm{mm}) + 0.57\times Path\ Diameter\ (25\ \mathrm{mm}) + 1.45\times Grade\ (Moderate) + 3.65\times Grade\ (Poor) + 3.89\times pN\ (1) + 4.46\times pN\ (2)$$

with p values of 0.265 for $MRI\ Diameter\ (10\ mm)$, 0.971 for $MRI\ Diameter\ (10\ mm)$, 0.455 for $Midline\ (Abuts)$, 0.026 for $Midline\ (Crosses)$, 0.003 for $pN\ (1)$ and <0.001 for $pN\ (2)$.

6.5.9 Results: post-surgical artificial neural networks

6.5.9.1 ACS

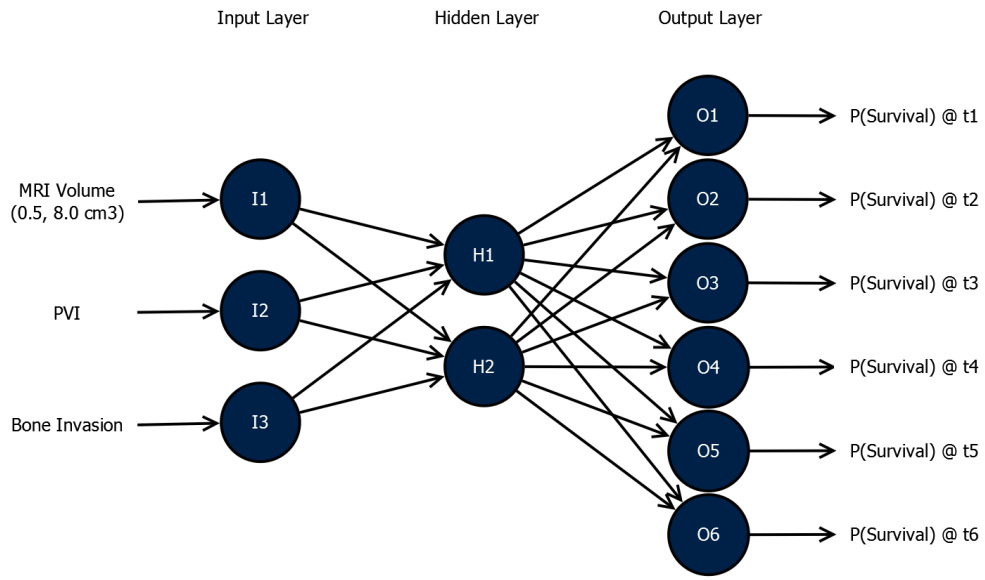


Figure 6.22: Structure of the top-ranked ANN for post-surgical prediction of 2-year all-cause survival. The explanatory variables were passed a two hidden nodes. The decay constant was 0.005. Each node in the output layer represents the survival probability at six equal time intervals {t1, t2, t3, t4, t5, t6}.

6.5.9.2 DRS

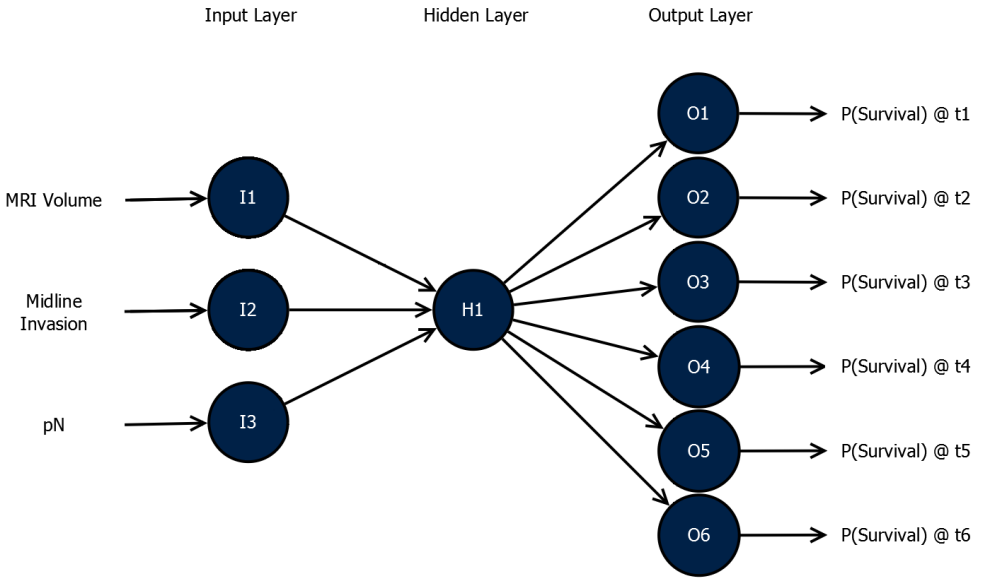


Figure 6.23: Structure of the top-ranked ANN for post-surgical prediction of 2-year disease-related survival. The explanatory variables were passed to one hidden node. The decay constant was 0.001. Each node in the output layer represents the survival probability at six equal time intervals {t1, t2, t3, t4, t5, t6}.

6.5.9.3 DFS

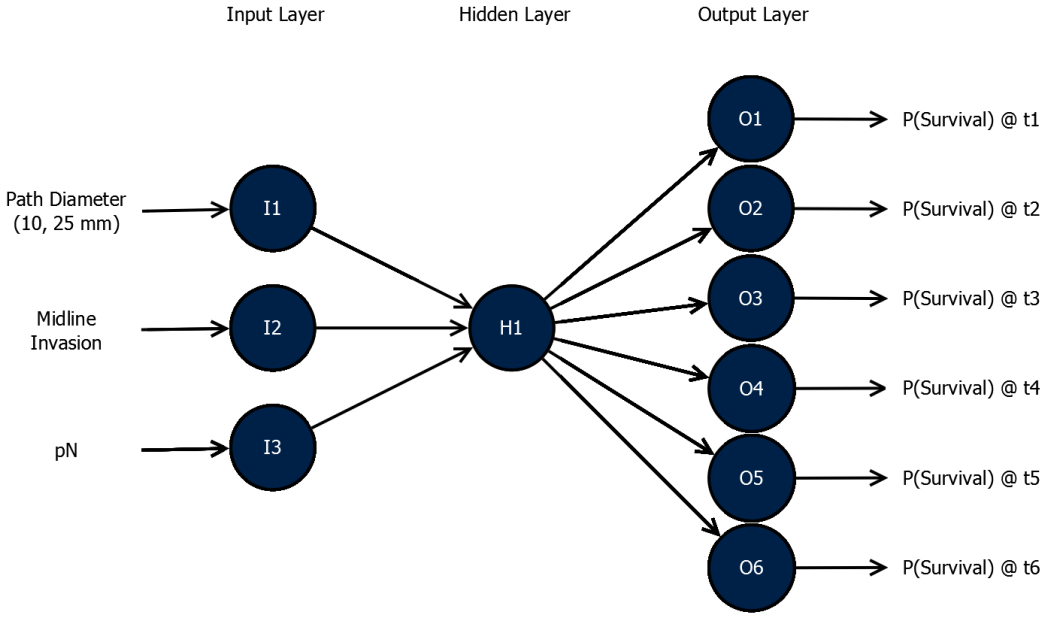


Figure 6.24: Structure of the top-ranked ANN for post-surgical prediction of 2-year disease-free survival. The explanatory variables were passed to one hidden node. The decay constant was 0.005. Each node in the output layer represents the survival probability at six equal time intervals {t1, t2, t3, t4, t5, t6}.

6.5.10 Results: post-surgical recursive partitioning

6.5.10.1 ACS

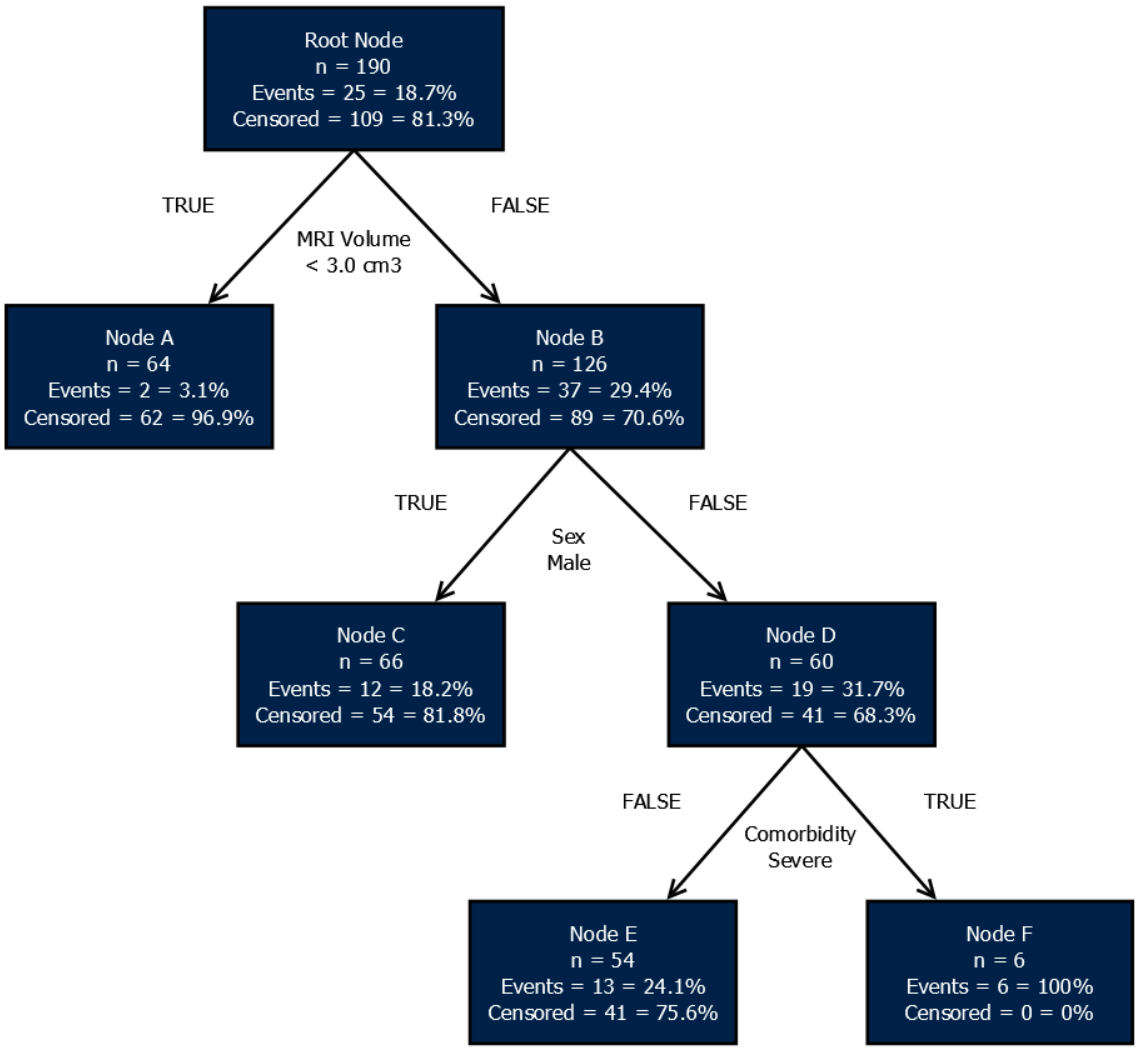


Figure 6.25: Structure of the top-ranked decision tree for post-surgical prediction of 2-year all-cause survival. The final tree consisted of 3 levels and had 4 terminal nodes. The complexity parameter was 0.01.

6.5.10.2 DRS

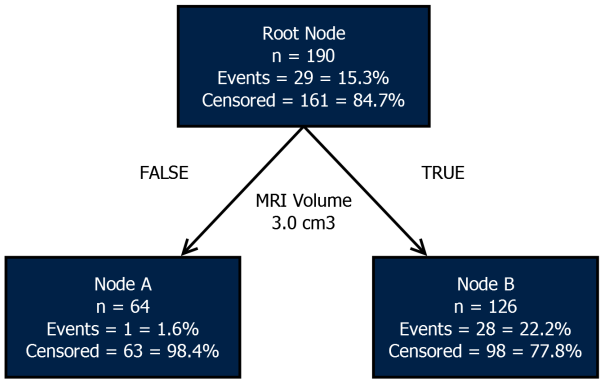


Figure 6.26: Structure of the top-ranked decision tree for post-surgical prediction of 2-year disease-related survival. The final tree consisted of 1 level and had 2 terminal nodes. The same result was obtained for complexity parameters of 0.05 and 0.1.

6.5.10.3 DFS

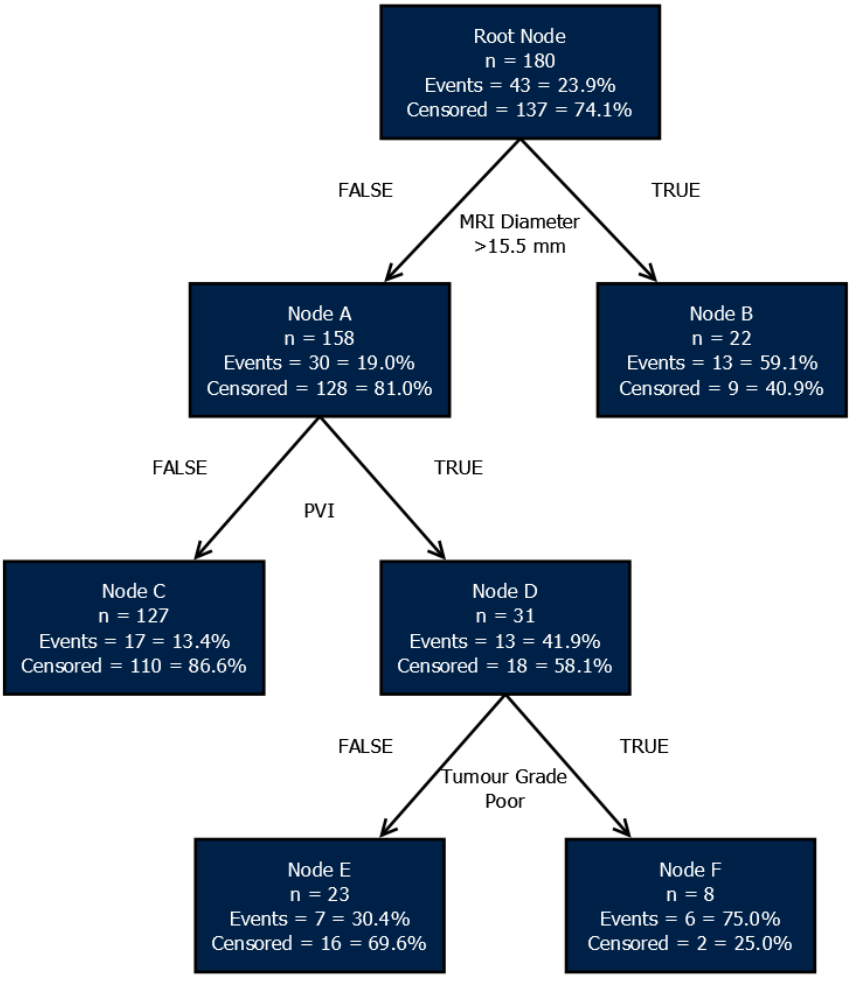


Figure 6.27: Structure of the top-ranked decision tree for post-surgical prediction of 2-year disease-free survival. The final tree consisted of 1 level and had 2 terminal nodes. The same result was obtained for complexity parameters of 0.05 and 0.1.

6.5.11 Results: post-surgical model summaries

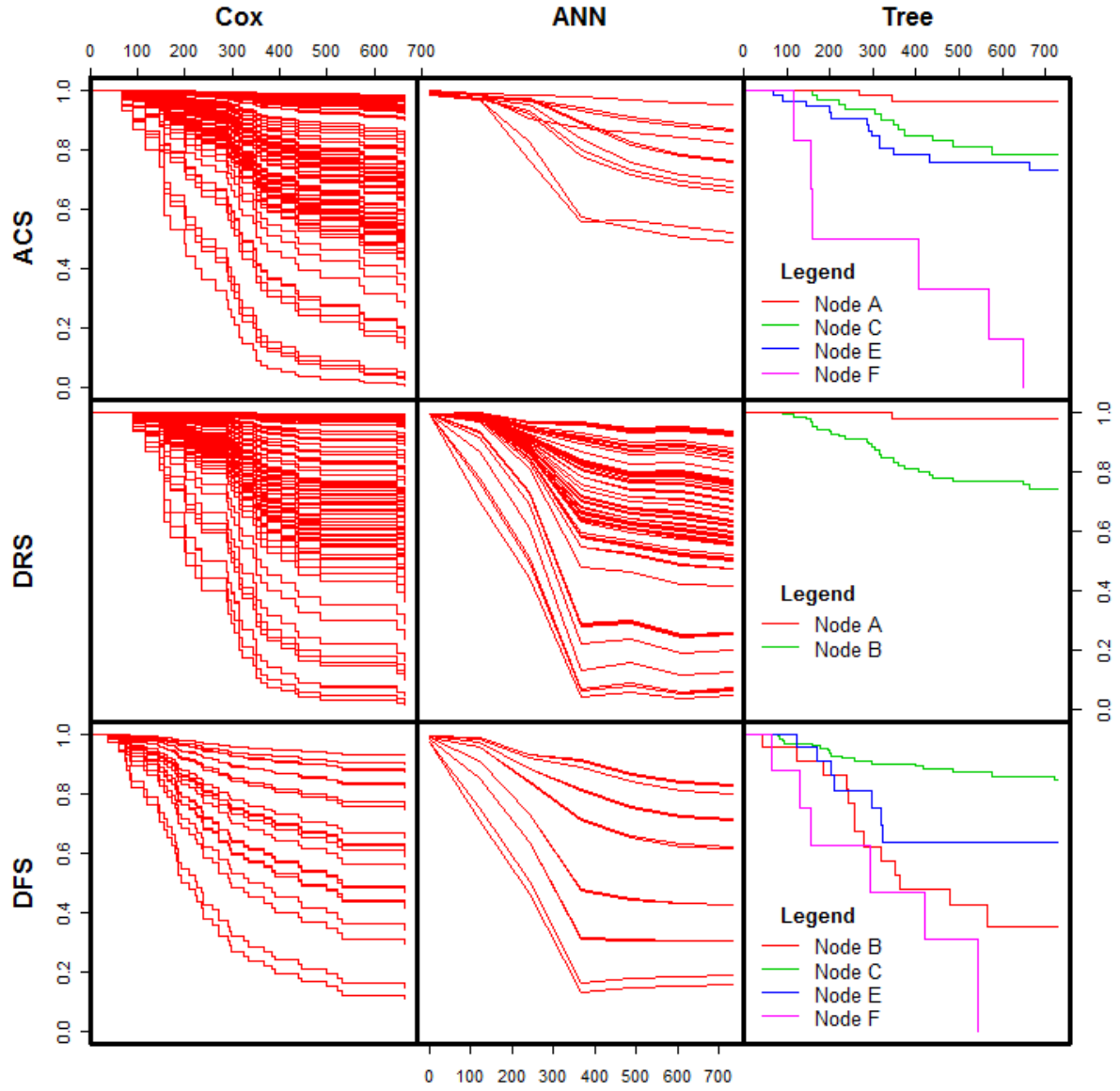


Figure 6.28: Plot of 2-year survival curves for the three post-surgical modeling methods. The first column is constructed from the post-surgical Cox model using the "survfit" function of R-Project "Survival" package. The function estimates the baseline hazard of the sample then uses the hazard ratio of each variable combination to plot the estimated survival curves. The middle column is constructed from the post-surgical artificial neural network output. The model output estimates the survival probability at six equal time intervals. In the plot, each survival curve represents a distinct combination of values of the three explanatory variables. The right column is constructed from the post-surgical recursive partitioning output. The cases were categorized according to the decision tree and the Kaplan-Meier curve plotted for each category. The vertical axis represents cumulative survival while the horizontal axis represents survival time in days. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival; Cox: Cox proportional hazards; ANN: artificial neural network; tree; recursive partitioning decision tree.

6.5.12 Discussion: post-surgical prediction

The results of post-surgical survival prediction were similar to that of pre-surgical modeling. The MRI-based volume was once again included in each of the top-ranked Cox and ANN models for ACS and DRS. Also like pre-surgical prediction, pathological diameter was the most valuable predictor of DFS.

Also mirroring pre-surgical survival prediction, pN factored prominently in all of the top-ranked ACS, DRS and RFS models for the Cox and ANN methods, except for the ACS ANN, which included LVI and bone invasion. Midline invasion was also included in all models with the exception of the ANN model of ACS and the Cox model of DFS, where it was replaced by tumour grade.

The result of recursive partitioning again differed markedly from the results of Cox and ANN modeling. Even within recursive partitioning, there was little similarity between ACS, DRS and RFS models. pN did not feature in any of the models with the exception of the second ranked ACS model, which also included path diameter (10, 25 mm) and midline invasion.

For ACS, volume (3.0 cm^3), sex and ACE-27 comorbidity were all included in the final tree. This model makes clinical sense as sex and comorbidity are documented determinants of mortality in the general population. However, the exclusion of pN is, again, suspicious.

The contribution of pN was also not reflected in the DRS decision tree. In fact, twenty different combinations of explanatory variables had the same cross-validated Briar Score and were considered to be the "top-ranked" model. Further analysis revealed that the trees had only one split, based on the measurement variable. However, one would expect all twenty of the measurement-based variables to be the same. Although, a mixture of diameter, volume and pT was present. This variability could indicate the inability of the recursive partitioning method to model DRS, however the fact that Briar Score was comparable to that of Cox and ANN modeling refutes this.

The top-ranked recursive partitioning model for DFS was based on pathological diameter (15.5 mm), LVI and tumour grade. Tumour grade was not included in the final decision tree. The tree

components are entirely plausible. Diameter reflects tumour extent and therefore the ability to achieve clear margins. LVI could be interpreted as a surrogate measure of tumour aggressiveness. As well, perivascular tumour extension can contribute to residual tumour following resection.

6.6 General discussion

The previous section presented the results and discussion for the modeling of OCLNM and survival in a compartmentalized fashion. It is important to bring the modeling results together in a general discussion in order to explore common themes and the implications to the natural history and prognosis of OCSCC. This chapter presents a unified discussion while also exploring the shortcomings of the multivariate modeling method.

6.6.1 MRI volume is a strong predictor of OCLNM

There is good evidence that MRI-based volume dichotomized at the median value of 3.0 cm^3 is a strong predictor of OCLNM. While the small sample size likely precludes the inclusion of multiple variables in prediction models, there is a strong possibility that LVI adds to post-surgical prediction.

The pre-surgical cross-validated NPV was greater than 95% for all three prediction methods. That is, patients classified as low risk by the models are at less than a 5% risk of harbouring occult metastases and may represent a group in which a watch-and-wait approach could be considered. Either of the three modeling methods would be suitable for this task, however the recursive partitioning model's ease of implementation might encourage acceptance.

The implications of these results are of key importance to informed decision making by both patients and surgeons. In the literature, it is often stated that a watch-and-wait approach to elective management of the neck may be justified when the probability of OCLNM is less than 20%. The actual choice is surgeon dependent. Many believe that 20% is an excessively high risk, given the effect of regional disease on survival, and choose elective neck dissection for all patients. The results of this study showed that only 5% patients with tumour volumes less than 3.0 cm^3 will had neck

disease on pathology. It may be that surgeons will find this level of risk much more satisfactory when contemplating elective neck dissection. Similarly, patients will be able to make an informed decision when they weigh the risk of OCLNM with the cosmetic and functional implications of neck dissection.

The results of pre-surgical modeling suggest that MRI-based volume and thickness may be interchangeable when used for OCLNM prediction. However, the results of post-surgical modeling overwhelmingly favour volume. It may be that volume would prove superior if the sample size was more robust and complex variable relationships better delineated.

The caveat of post-surgical prediction of OCLNM is that it is irrelevant when neck dissection has been performed. However, in situations in which neck dissection is not performed with the primary resection this information might be of some value. As well, this result indicates that there is some value to evaluating LVI on pre-surgical biopsy, which is currently not widely available.

6.6.2 Volume-based models best predict survival

With the exception of DFS prediction, there was remarkable consistency across the board. Pre-surgically, the top-ranked Cox and ANN models for ACS and DRS were all comprised of MRI volume and cN. All of these models also included midline invasion, with the exception of the ANN model for DRS, which included bone invasion. Post-surgically, volume remained the measurement component of all top-ranked Cox and ANN models for ACS and DRS. The remainder of each of these models were composed of midline invasion and pN. The one exception was the ACS ANN, which used PVI and bone invasion. However, the third ranked model, with a Brier score only 0.0003 higher than the top-ranked model, consisted of volume, midline invasion and pN.

As predicted, CLNM status was an important component of the top-ranked prognostic models for ACS and DRS. As well, midline invasion has the potential to act as an anatomically-based predictive factor that is superior to bone or extrinsic muscle invasion for survival prediction.

The results for DFS are less clear cut. Volume was the measurement component of the top-ranked

Cox model, while diameter took this role in the ANN model. The ANN model was also the overall best DFS model. The ANN model also included midline invasion and cN.

There is a case for a unified model for pre-surgical ACS and DRS prognosis using an artificial neural network with the inputs MRI volume, midline invasion and cN. Similarly, there is a case for a unified post-surgical model for ACS and DRS artificial neural network with the inputs MRI volume , midline invasion and pN.

Volume was superior to all other measurement types for the prediction of ACS and DRS. This is not surprising given noted shortcomings of maximum diameter as a TNM criterion. However, given the body of literature espousing depth of invasion as a candidate for a revised TNM staging protocol., it's minimal presence in any of the top-ranked models is notable. There is a novel element to this result as this is one of few works to demonstrate the utility of MRI-measured volume as a determinant of prognosis.

6.6.3 Artificial neural networks are superior predictors of survival

ANN-based models were superior to Cox proportional hazards-based models for ACS, DRS and DFS prediction. There are no publications in which an ANN has been documented to be superior to the more conventional Cox proportional hazards method of survival analysis for OCSCC. There was also a general trend towards lower Briar Scores from post-surgical models compared to pre-surgical models. This is expected since pathological analysis is considered to be the gold standard for the evaluation of most tumour characteristics.

Using the ANN model it is possible to provide a probability of two-year ACS and DRS for each individual patient based on the tumour volume, midline invasion and CNLM status. This individualized approach is an emergent theme in disease prognostication, but it is focused mostly on molecular markers. The ANN model can offer an individualized approach using readily available information and can facilitate informed decision making by both patient and surgeon.

6.6.4 Shortcomings

It is important to discuss the shortcomings of the modeling methods explored in this Chapter. Problems and shortcomings were related to variable selection and pre-processing, cross-validation, computer processing, error measurement and the selection of the optimum model.

6.6.4.1 Variable selection

Variables were selected for inclusion in multivariate models according to their ability to predict OCLNM or survival on univariate analysis. While this was a necessary step to limit the number of variables, there is a possibility that important factors were excluded. Situations could exist in which variables are of little value in isolation, but when considered in conjunction with another explanatory variable able to predict the outcome. Ideally, variable combinations could be screened, but such a process quickly becomes prohibitive with large numbers of variables.

6.6.4.2 Pre-processing

For continuous variables, statistical transformation is sometime a necessary step prior to inclusion in a multivariate model. This is especially true for regression models, which assume that there is a linear relationship between the explanatory variable and the outcome. Continuous variables can be evaluated for higher order interactions. However, evaluation of all higher order relationships is prohibitive and was not performed. It may be that the discovery of higher order relationships among the continuous variables would have improved the results regression modeling.

Similarly, the MRI-derived measurement data was bimodal and skewed to the right as a result of the designation of tumours invisible on MRI as 0 cm^3 or 0 mm. Log transformation did not correct the skew. Categorization into groups using the mean and quartiles overcame the distribution problem, but resulted in loss of information. Because it was unclear which method was best, native, log transformed and categorized data were all included in model building. It may have been more appropriate to choose the most suitable method prior to model building.

6.6.4.3 Cross-validation

10-fold cross-validation was carried out to assess the ability of each model to generalize. Ideally, trained models are tested on external datasets. However, such a wealth of cases was unavailable. Bootstrapped cross-validation might have been a more appropriate method as it can be repeated a large number of times and the results averaged. 10 fold cross-validation, by contrast, uses only 10 test sets and can be prone to sampling error. However, bootstrapping is computationally expensive and would have pushed computation times beyond reasonable limits.

6.6.4.4 Computer resources required

Many of the model building processes are computationally intensive. With large number of variables, computation times can become unreasonable. In this work, execution of the R-project script for building post-surgical recursive partitioning models took in excess of 17 hours. By contrast, the process took 8 hours for ANN modeling and 3 hours for Cox modeling. An estimate of the total time to build all of the models evaluated in this study is in excess of 100 hours.

This time depends on available computer resources. All R-scripts were custom written by the author with a basic knowledge of computer programming. There is probably scope to optimize the R-scripts resulting in improved execution times.

6.6.4.5 Briar Score

The method used to estimate Briar Score was based on measurement of the probability of survival at a single point in time, two years after surgical resection. A more valid measure might be to evaluate the similarity of entire survival curves from time zero to the end point at two years. Other methods to evaluate the ability of survival models to generalize do exist, such as the survival ROC (receiver operating characteristic). Unfortunately, a complete discussion is beyond the scope of this work.

6.6.4.6 Model suitability

Volumes of research exist that are devoted to Cox proportional hazards, ANNs and recursive partitioning analysis. A complete discussion of the breadth of the literature is beyond the scope of

this work; however the reader is reminded that there are many variations of the three methods. It may be that these variations would prove more adept at predicting OCLNM and survival. One of the goals of this work was to evaluate clinical prediction methods that are traditionally not available to the clinical researcher. In doing so, the goal is to expand the options available to the OCLNM researcher and lay the foundation for further refinement of the experimental technique.

6.7 Conclusions

In this Chapter patient and tumour factors, identified as predictors of OCLNM and survival in Chapter 5, were used to create multivariate regression, artificial neural network and recursive partitioning models. These models were assessed for their ability to predict OCLNM and survival, with emphasis on generalization to external data. This is the only OCSCC study to attempt to use recursive partitioning and ANNs to predict OCLNM and survival.

For OCLNM prediction, it was not possible to create complex predictive models with multiple explanatory factors due to the small sample size. However, it is clear that the volume of tumour invasion is of key importance.

For survival, ANNs proved superior to both Cox models and recursive partitioning. There is a case for the creation of a unified ANN model that uses volume, N-stage and midline invasion to predict ACS, and DRS.

Further work is required to expand the sample size and develop more complex models of OCLNM and survival. Doing so will enhance predictive ability and further allow surgeons to offer patients individualized predictions of OCLNM and prognosis.

7 The extrinsic muscles of the tongue

7.1 Introduction

The TNM staging criteria of the American Joint Committee on Cancer specify that invasion of the extrinsic muscles of the tongue, including the styloglossus, hyoglossus, genioglossus and palatoglossus, classify a tumour as T4.

However, the styloglossus and hyoglossus lie in the lateral tongue, the oral cavity subsite most often involved by SCC, and are very superficial. The invasion of these lateral extrinsic muscles is largely ignored by clinicians when clinical staging takes place. This is due to either lack of knowledge of their superficial position, or an experience-based assessment that their invasion is not clinically significant.

This Chapter seeks first to evaluate the Visible Human Female as a resource for defining the normal anatomy of the oral cavity and tongue. The position, course and relationships of the extrinsic and intrinsic muscles of the tongue are compared to classical descriptions. Second, detailed measurements of the distance of the styloglossus and hyoglossus from the lingual surface are taken to provide objective data on the superficial position of these muscles.

7.1.1 Chapter goals

The goals of this chapter are:

1. *To review the role of extrinsic muscle invasion in TNM staging of OCSCC*
2. *To review the anatomy of the oral tongue, including the extrinsic and intrinsic muscles.*
3. *To highlight the Visible Human Project and the role of the Visible Human Female in studying lingual anatomy.*
4. *To use the VHF to review and define the anatomy of the oral tongue and to compare the result to classical anatomic descriptions.*
5. *To use the VHF to measure the distance from the lingual surface to the superficial border of the lateral extrinsic tongue muscles.*

7.1.2 Organization of sections

The chapter begins with a description of the current T4 staging criteria for the oral cavity, focusing on the role of the extrinsic muscles. The anatomy of the lingual myoarchitecture is outlined and the specific muscle groups discussed. The Visible Human Project is introduced as well as its role in anatomical description.

Next, the methodology is described including the selection and pre-processing of VHF images, the identification and description of the intrinsic and extrinsic muscles of the tongue and measurement of the distance from the lingual surface of the styloglossus and the genioglossus.

The methods are followed by the results and discussion. The validity of the VHF as a means for anatomic study is outlined. The implications of the measurement study for interpretation of TNM staging criteria are described. The chapter ends with a brief summary and conclusions.

7.2 Background

The anatomy of the oral cavity was explored in Chapter 2 but lacked a detailed description of the lingual musculature. This section begins by outlining the basic anatomy and function of the extrinsic muscles of the tongue and concludes with a description of the Visible Human Project.

7.2.1 Definition of a T4 tumour of the oral cavity

The T-staging criteria of the current TNM system for the oral cavity are largely measurement based for T1 to T3. However, the T4 classification is based on the invasion of various critical structures. T4a and T4b for defined by the AJCC as follows [222]:

> *T4a (oral cavity): Tumor invades adjacent structures (e.g., through cortical bone, into deep [extrinsic] muscle of tongue [genioglossus, hyoglossus, palatoglossus, and styloglossus], maxillary sinus, skin of face).*
>
> *T4b: Tumor invades masticator space, pterygoid plates, or skull base and/or encases internal carotid artery.*

In this capacity, the subjective term "deep" is a misnomer. Based on classical anatomic descriptions, components of the styloglossus, hyoglossus and palatoglossus run superficially. Even the genioglossus, described as deep throughout most of its course, is covered by only a thin layer of mucosa in the anterior floor of the mouth. "Deep" is a descriptor that is highly dependent on position in the oral cavity.

7.2.2 The lingual myoarchitecture

The myoarchitecture of the tongue was first reported by Abd-El-Malek in 1939 [223], and expanded by the work of Miyawaki in 1974 [224]. The result of their work is the classical description of lingual muscle anatomy found in modern texts.

The muscles of the tongue are divided into two groups: the extrinsic and the intrinsic muscles. While these muscles are often attributed discrete roles in tongue movement. This is an oversimplification; the reality is a complex synergism between the components. The tongue has been described as a muscular hydrostat, a muscular organ that maintains constant volume during movement [225]. Examples include elephant trunks and squid tentacles. Nevertheless, it is useful to consider the component parts for the purpose of illustration.

7.2.2.1 The extrinsic muscles

The extrinsic muscles are anchored to bone and connective tissue that lie external to the lingual bulk. The muscle insertions do not consist of discrete structures, but are a result of the muscle fibre intermingling with the intrinsic muscles and other extrinsic muscles. Because of their external anchoring, the extrinsic muscles primarily move the tongue in relation to the oral cavity but also contribute to changes in shape. The extrinsic muscles and surface landmarks are diagrammed in Figure 7.1.

The genioglossus

The genioglossus is attached by a short tendon to the superior mental spine of the mandible. Its fibres enter the tongue inferiorly and attach to the dorsum. Its fibres also insert into the body of the

hyoid bone. Acting bilaterally, the genioglossus muscles depress the tongue and create a central concavity. Acting unilaterally, they ''wag'' the tongue to the contralateral side. They are also involved in protruding the tongue and retracting the apex when the tongue has been protruded [2].

The hyoglossus

The thin, rhomboid shaped hyoglossus originates from the body and greater horn of the hyoid, and inserts into the inferolateral aspect of the tongue. It depresses the tongue, makes the surface convex and assists in retraction [2].

The styloglossus

The small, thin styloglossus originates from the anterior styloid process and stylohyoid ligament and passes inferoanteriorally, inserting into the inferior tongue and interdigitating with the hyoglossus. The styloglossus retracts the tongue and creates a central concavity during deglutition [2].

The palatoglossus

The palatoglossus originates in the aponeurosis of the soft palate and inserts into the lateral tongue, where it becomes intertwined with the intrinsic transverse muscles. The palatoglossus elevates the tongue and lowers the soft palate during deglutition [2].

7.2.2.2 The intrinsic muscles

The intrinsic muscles entirely within the substance of the tongue and have no external attachments. They primarily alter the shape of the tongue but also play a role in movement. The intrinsic muscles are shown in Figure 7.1.

The superior longitudinalis

The superior longitudinal muscle runs from the root to the apex on the dorsal surface, deep to the mucous membrane. It curls the apex of the tongue, pointing it towards the nose.

The inferior longitudinalis

The inferior longitudinal muscle runs from the root to the apex along the inferior surface of the

tongue. It curls the tip of the tongue towards the chin. Acting with the superior longitudinal muscle, it retracts the protruded tongue, and makes its shape short and thick [2].

The transversus and verticalis

The transverse muscle fibres lie deep to the superior longitudinal muscles, arising from the midline septum of the tongue and running laterally. Their role is to narrow and elongate the tongue. The vertical muscles run inferolaterally from the dorsum. They flatten and broaden the tongue. The transverse and vertical muscles interdigitate and, acting together, elongate and narrow the tongue, aiding protrusion [2].

7.2.3 The visible human project

The Visible Human Project (VHP) is an initiative of the US National Library of Medicine. Two whole human cadavers, the male (VHM) in 1994 and the female (VHF) in 1995, were frozen and sectioned. A high-resolution photograph was taken of each section. The cadavers were also imaged by CT and MRI prior to being sectioned. The ultimate goal of the VHP was to create detailed, three-dimensional datasets of the complete human anatomy, available free of cost to the scientific community [226].

The VHF is the more detailed of the two datasets, with an in-slice resolution of 2048 X 1216 pixels and pixel dimensions of 0.33 X 0.33 mm, with 24 bits of color. Slice thickness was 0.33 mm to create isometric voxels. A total of 5189 images, comprise a database of approximately 40 GB [226]. Examples of images produced by the VHP are shown in Figure 7.2.

7.2.4 The usefulness of the VHF

The VHF represents a source of detailed lingual anatomy. The thin slice thickness, combined with high resolution photography, makes the information content of the dataset unparalleled.

With the dataset arranged as a series of images in a stack, the viewer is able to navigate from slice to slice in a manner familiar to those with experience viewing CT or MRI. The isometric dimensions of the VHF voxels make the data ideal for multiplanar reformatting (MRP) and 3D rendering.

Figure 7.1: The muscular anatomy of the tongue. The genioglossus, hyoglossus, styloglossus and palatoglossus comprise the extrinsic muscles of the tongue. The superior longitudinalis, inferior longitudinalis, transversus and verticalis muscles comprise the intrinsic muscles of the tongue. The four extrinsic muscles of the tongue originate outside of the tongue and attach to it. Their primary role is to move the tongue. Source: [2]

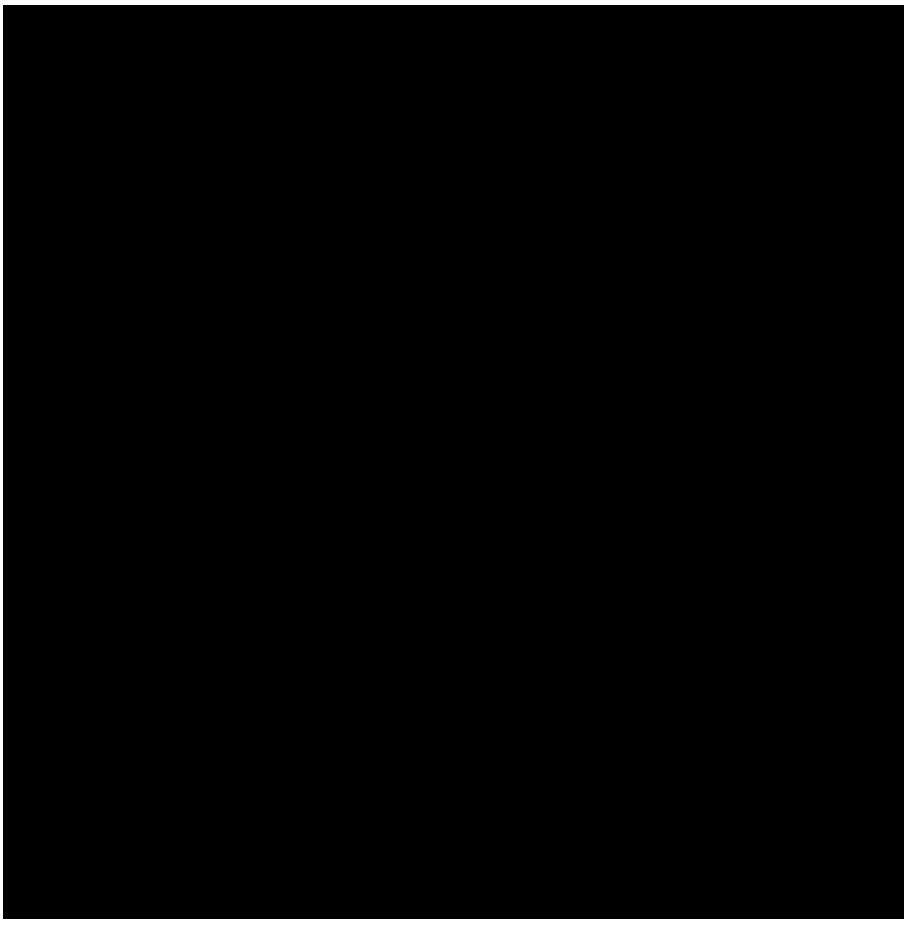


Figure 7.2: The Visible Human Male. On the left, three-dimensional rendering of the VHM. Anterior and right lateral views are shown. The VHP was "sliced" and photographed as four sections, rather than a complete cadaver, thus the separations seen below pectoralis major, the upper thigh and below the knee. Axial and coronal view of the VHF. At top right, an axial slice of the VHF is shown at the level of the mandibular ramus. The paired genioglossi are clearly shown. Bottom right, a coronal slice of the VHF produced by a multiplanar reformat of the axial data. Again, the paired genioglossi are seen as well as other aspects of lingual anatomy. Source: [226]

The VHF offers surgeons and radiologists the opportunity to view the oral cavity and tongue *in situ.* This is not possible with resected specimens. Comparison with classical descriptions of lingual anatomy is a first step towards validation which, at the same time, may provide new insights. Of particular interest is the positioning of the intrinsic and extrinsic muscles relative to the surface, especially in the lateral tongue.

7.3 Methods

This section consists of three parts. The first part describes the sources and required processing for the images used in this work. This is followed by procedures for identification of the extrinsic musculature and distance measurement using the VHF.

7.3.1 Source images and pre-processing

Axial and coronal images of the VHF were downloaded as JPEGs from the Visual Human Browser website [227]. The images were imported into Adobe Lightroom 2 (V. 2.5, Adobe Systems, San Jose, CA) and cropped to include only structures necessary for this investigation, and also to reduce image size.

7.3.2 Identification of lingual musculature

The cropped images were imported into ImageJ as image stacks, allowing for navigation and visualization similar to conventional CT and MRI viewing software. The extrinsic muscles were identified and followed from bony insertion to lingual insertion. The appearance, position and anatomic relationships of the extrinsic and intrinsic muscles were described and compared to classical descriptions.

7.3.3 Measurement of distance from surface mucosa

Measurements were taken using the ImageJ measurement tool calibrated to the 0.33 X 0.33 mm pixel dimensions and exported to Excel for evaluation.

The distance relative to the lingual surface was measured for the styloglossus and hyoglossus. It was

obvious that this distance was least in the lateral dimension, thus measurements were limited to this area. The distance was defined as the length of a line drawn perpendicular to the mucosal surface to the most superficial muscle fibres. Distances were not measured for the deep lying genioglossus or the poorly seen palatoglossus. Measurements were taken on both axial and coronal VHF images.

In cases in which the lingual mucosa was juxtaposed to, and indistinguishable from, the surface mucosa of other oral structures (for example in the lingual sulci), the combined thickness of the two mucosal surfaces was used. This was necessary to avoid choosing an arbitrary separation point between the two surfaces. In each slice, a maximum of five measurements were taken in the lateral tongue. Measurement location was based on the operator's perception of smallest distance.

Measurements were taken on every fifth axial and coronal slice, reducing the slices processed by 80%. This was appropriate given the fine (0.33 mm) slicing of the VHF. An example of the distance from surface measurements is shown in Figure 7.3.

7.4 Results

This section summarizes the results of anatomic description and distance measurement described in the methods.

7.4.1 Description of the extrinsic muscles as seen on the VHP female

7.4.1.1 The styloglossus

The styloglossus is seen to arise on the anteromedial surface of the distal styloid process, at the same level as the origin of the stylohyoid and inferior to the origin of the stylopharyngeus. The muscle descends anteromedially, its axis oblique to the axial plane, passing between the superior and middle pharyngeal constrictor muscles. At the lateral tongue base, the palatoglossus merges with both the intrinsic muscles (particularly transversus and superior longitudinalis) as well as the styloglossus.

Upon reaching the base of the tongue, the muscle's directional axis enters the axial plane and appears to split into posteromedial and anteromedial groups of fibres. Those fibres passing

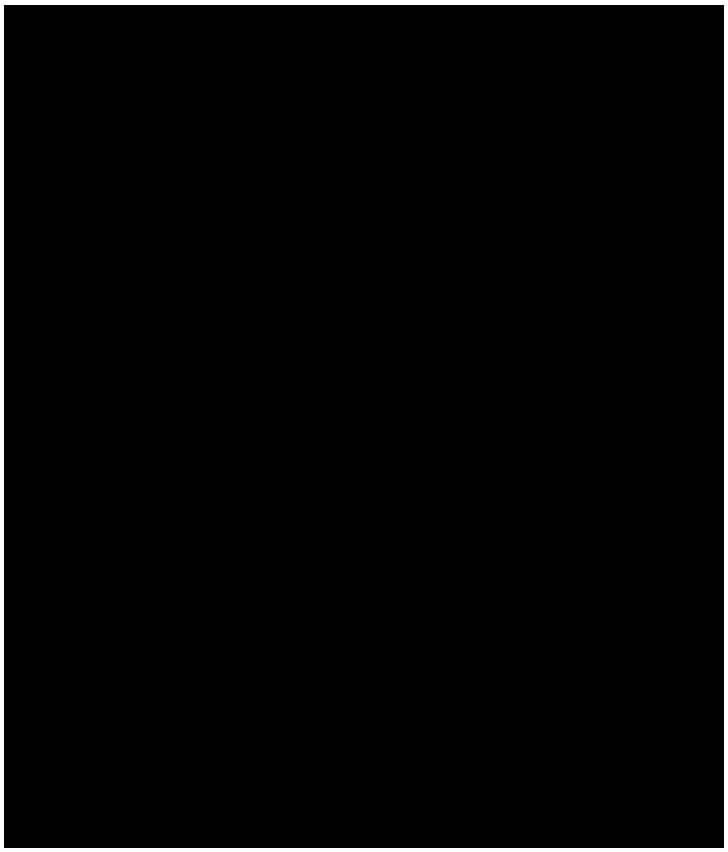


Figure 7.3: Measurement of extrinsic muscle distance. Five measurements, indicated by green lines, of the distance of the left hyoglossus are shown. A zoomed coronal view of the VHF is shown, located approximately in the middle one-third of the tongue. In this case, the mucosa of the lateral tongue and medial alveolar ridge are juxtaposed. Measurements were taken using the combined width of the mucosal layers.

posteromedially appear to merge with the vertically oriented fibres of the hyoglossus muscle and the superior longitudinalis. The anteromedial fibres of the styloglossus follow the inferolateral tongue from the posterior margin of the hyoglossus to within a few centimeters of the apex, at which point the fibres interdigitate with, and are indistinguishable from, the inferior longitudinalis.

Throughout its course from the tongue base to apex, the styloglossus remains superficial and distinct from the hyoglossus. It is distinguishable by its relatively lighter shade of red and by its muscle fiber orientation. A thin but visible fat plane separates the muscles throughout their juxtaposition. Coronal and axial views of the styloglossus are described visually in Figure 7.4 and Figure 7.5.

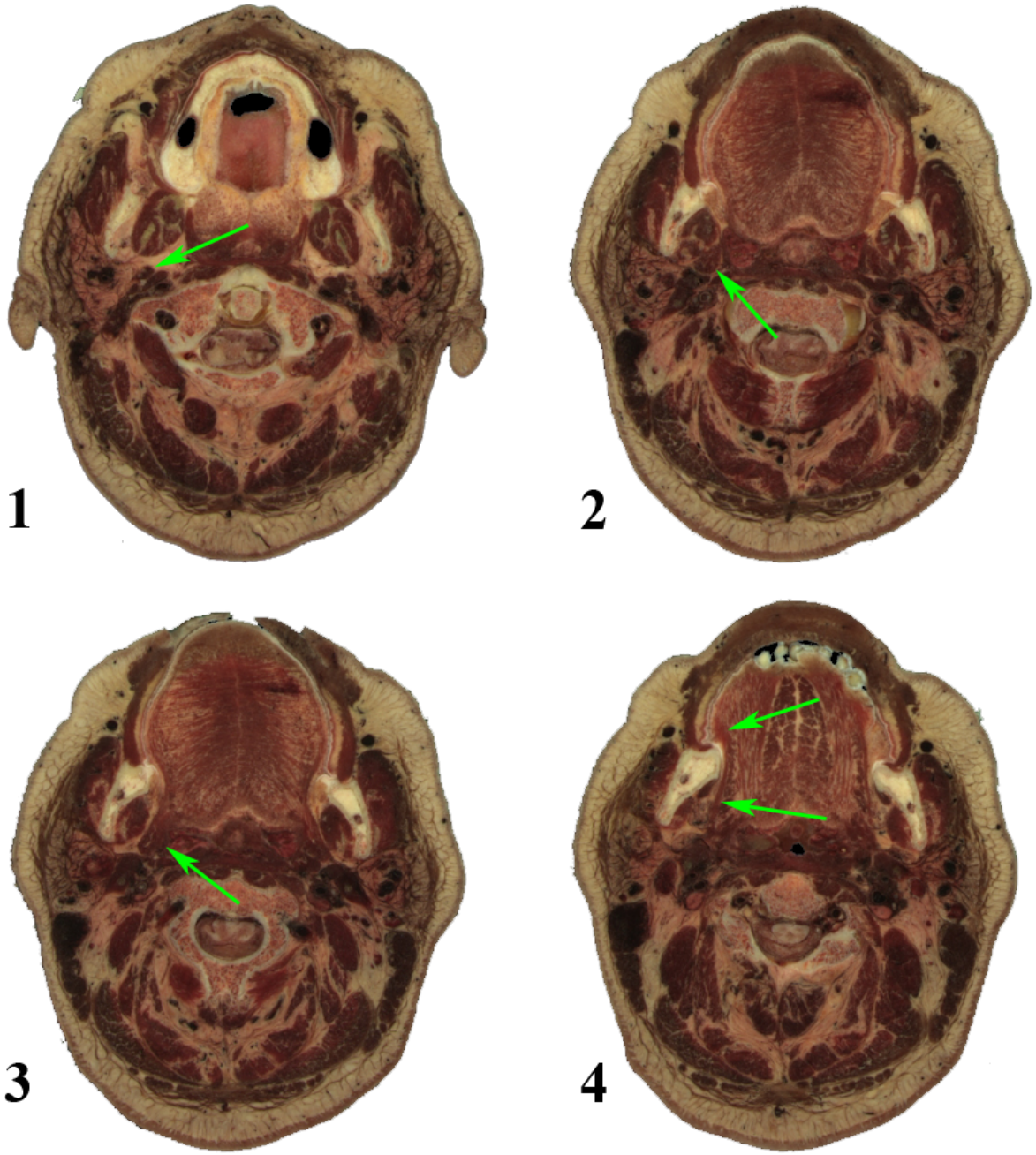


Figure 7.4: Axial VHF view of the styloglossus. The styloglossus is highlighted by green arrows. In image 1 the styloglossus is highlighted at its origin on the anteromedial surface of the distal styloid process. Image 2 shows the SG as is descends anteromedially, passing between the superior and middle constrictors. Image 3 shows the SG at its insertion into the tongue. Image 4 shows the fibres of the SG, oriented in the axial plane, as they pass just beneath the lateral surface of the tongue.

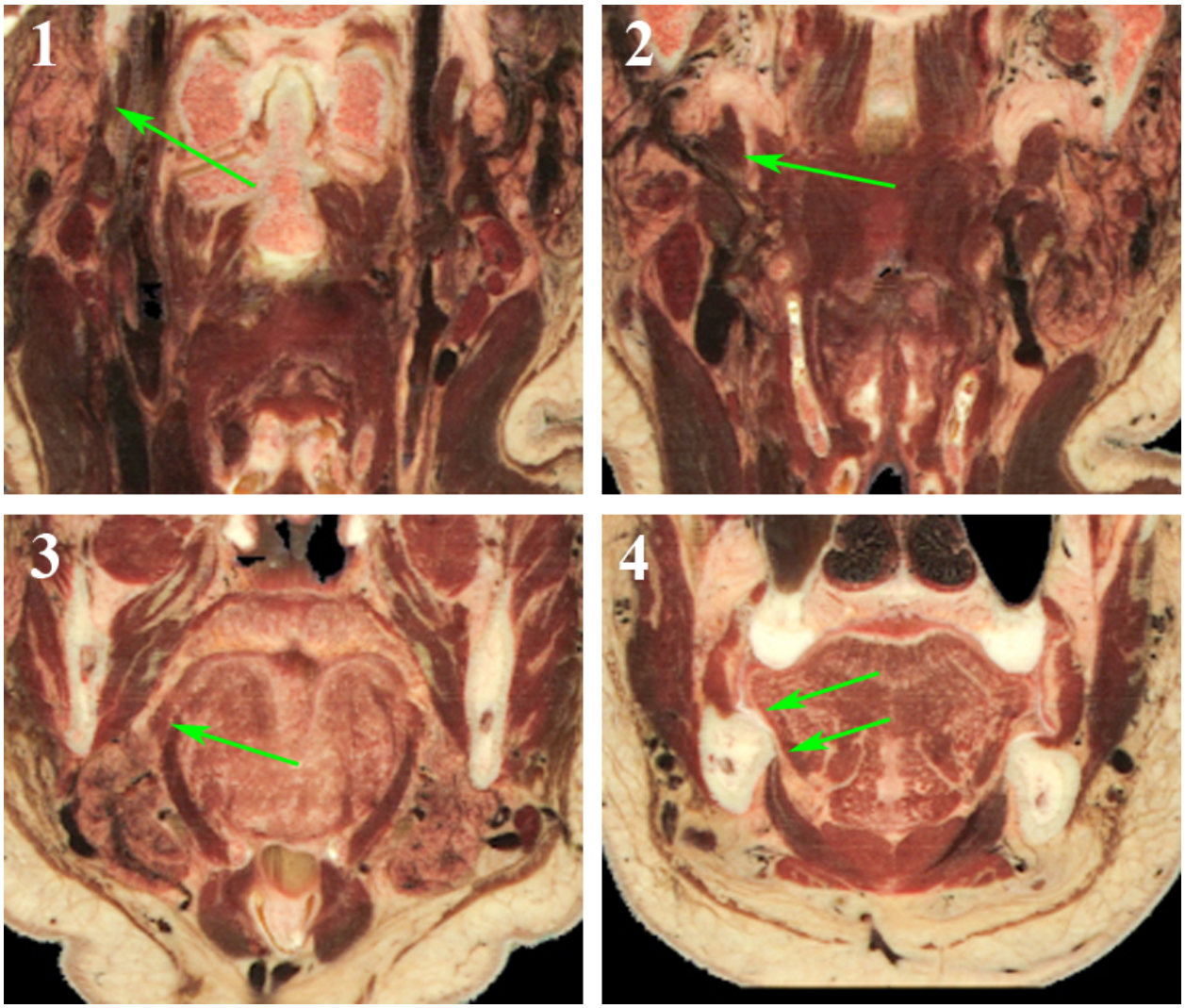


Figure 7.5: Coronal VHF view of styloglossus. The styloglossus is highlighted by green arrows. This series of images parallels those seen in Figure 7.4. The styloglossus is highlighted by green arrows. In image 1 the styloglossus is highlighted at its origin on the anteromedial surface of the distal styloid process. Image 2 shows the SG as is descends anteromedially, passing between the superior and middle constrictors. Image 3 shows the SG at its insertion into the tongue. Image 4 shows the fibres of the SG, oriented in the axial plane, as they pass just beneath the lateral surface of the tongue.

7.4.1.2 The hyoglossus

The hyoglossus arises from the greater cornu and lateral surface of the hyoid bone. The muscle then extends superolaterally and inserts into the lateral aspect of the posterior half of the tongue. The hyoglossus is rhomboid in shape with its cranial-caudal and anterior-posterior axes slightly skewed relative to the sagittal plane. In the posterior one-third to one-half of the tongue, the hyoglossus is seen to extend vertically from the hyoid bone to as much as 80% of the cranial-caudal lingual diameter. In the anterior tongue, the vertical extent of the hyoglossus is limited to the inferolateral tongue.

The hyoglossus is remarkable for its deep red color relative to the intrinsic muscles. Near its origin on the lateral surface of the hyoid bone the hyoglossus is uniform in appearance with little, if any, fibrofatty infiltration. However, as the muscle ascends, fibrofatty infiltration gradually becomes more pronounced, separating into muscle fiber groupings of ever-decreasing caliber. This gradual infiltration continues until the hyoglossus is no longer distinguishable from the intrinsic muscles.

As described above, the hyoglossus appears to be distinct from the styloglossus, except at its posterior edge where the muscles merge.

Coronal and axial views of the hyoglossus are described in Figure 7.6 and Figure 7.7.

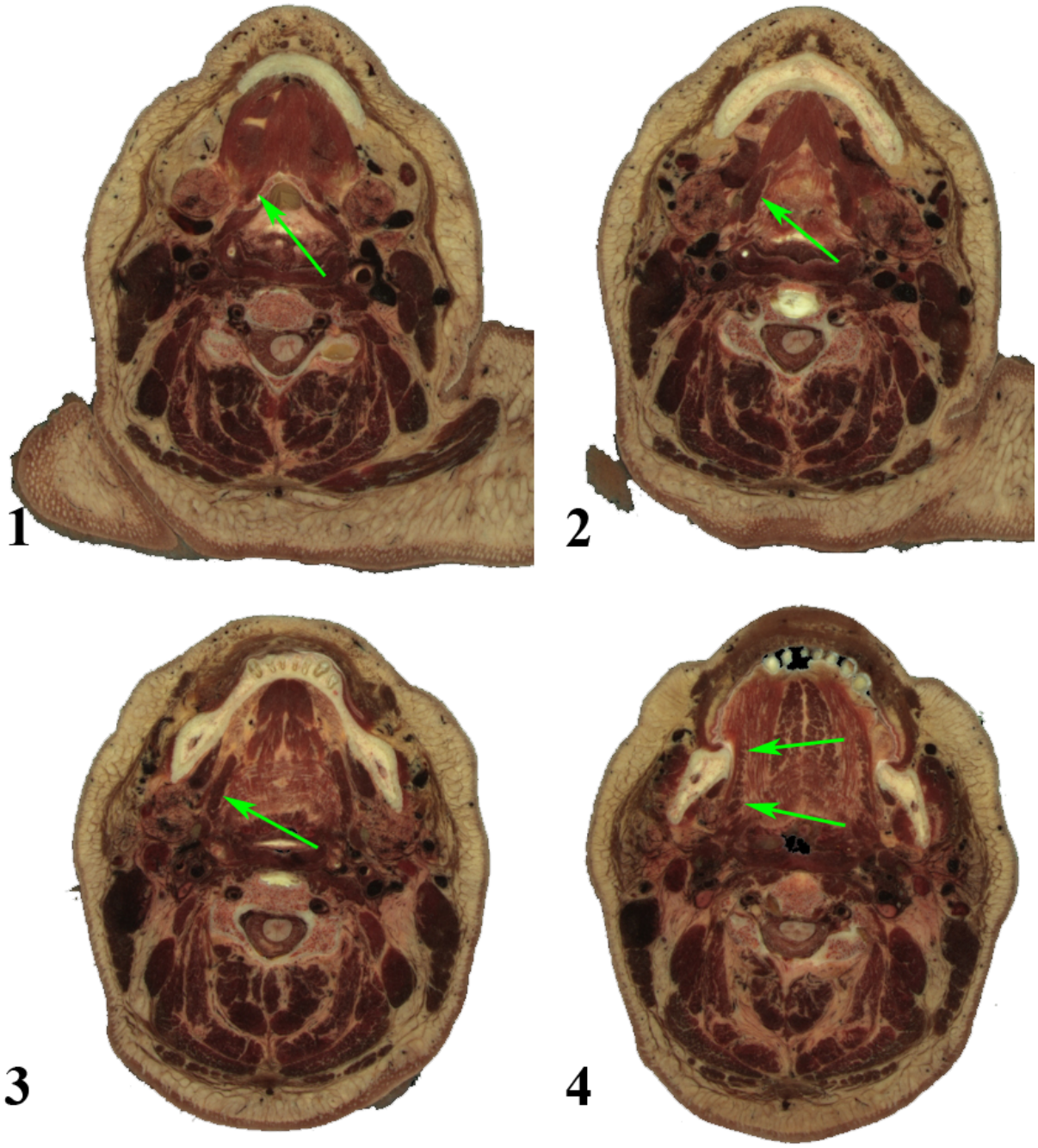


Figure 7.6: Axial VHF view of the hyoglossus. The hyoglossus is highlighted by green arrows. In image 1 the origin of the HG on the lateral surface of the hyoid bone is shown. Image 2 shows the body of the HG as it expands in AP diameter and moves superolaterally. Image 3 is at the level of the mandibular insertion of the GG and shows the HG assuming its rhomboid shape. In image 4, fibrofatty infiltration of the HG is evident as it begins to merge with the intrinsic muscles of the tongue.

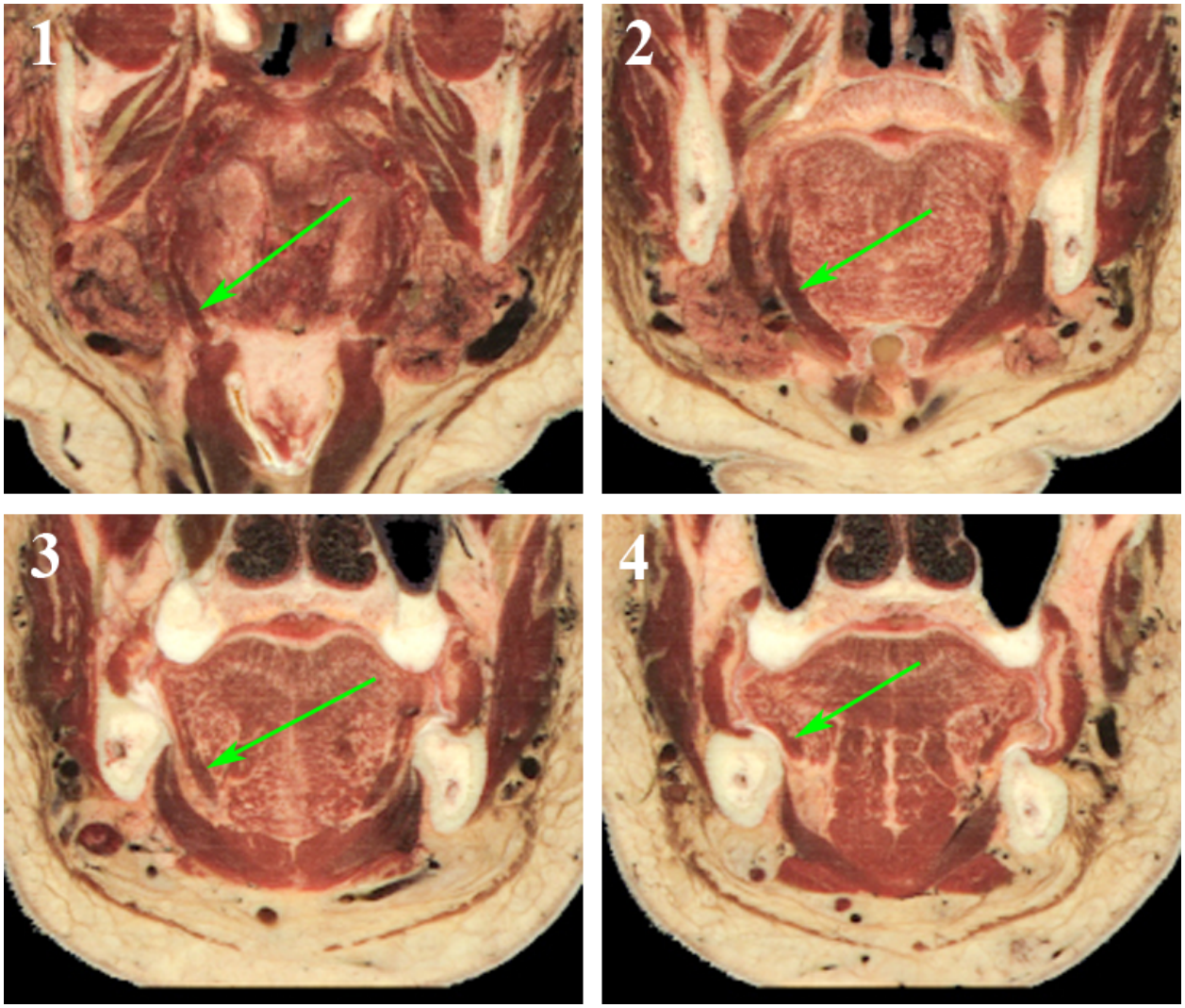


Figure 7.7: Coronal VHF view of the hyoglossus. The hyoglossus is highlighted by green arrows. This series parallels those seen in Figure 7.6. In image 1 the origin of the HG on the lateral surface of the hyoid bone is shown. Image 2 shows the body of the HG as it moves superolaterally. Image 3 demonstrates that as we move anteriorly in the tongue the HG become lesser in extent, a consequence of its rhomboid shape. In image 4, the HG is very close to the surface of the lateral anterior two-thirds of the tongue. Only the styloglossus and a thin layer of mucosa separate it from the surface.

7.4.1.3 The genioglossus

The genioglossus arises from the inferior mental spine on the lingual surface of the mandible at the symphisis menti. Its origin is immediately superior to that of the geniohyoid muscle. The genioglossus spreads from the origin in a fan-like pattern, well demonstrated on sagittal MPRs of the VHF, until it eventually inserts or, more appropriately, interdigitates with the transversus and verticalis muscles.

The genioglossus, like the hyoglossus, has a deep hue of red relative to the intrinsic muscles of the tongue. Also like hyoglossus, the composition of the genioglossus is relatively uniform near its origin but displays increasing amounts of fibrofatty infiltration as the distance from the origin increases. This infiltration separates the muscle into increasingly smaller fiber bundles until complete integration with the intrinsic muscles occurs.

In the midline, the fibrofatty median raphe of the tongue is well defined and separates the genioglossi into distinct left and right muscles. Laterally, the thin but distinct fibrofatty paramedian septum separates the genioglossus from the intrinsic muscles of the tongue. Both the median and paramedian septi become less pronounced as the genioglossus integrates into the intrinsic muscles and these are not found in the superior tongue.

Sagittal, coronal and axial views of the genioglossus are described in Figure 7.8 and Figure 7.9.

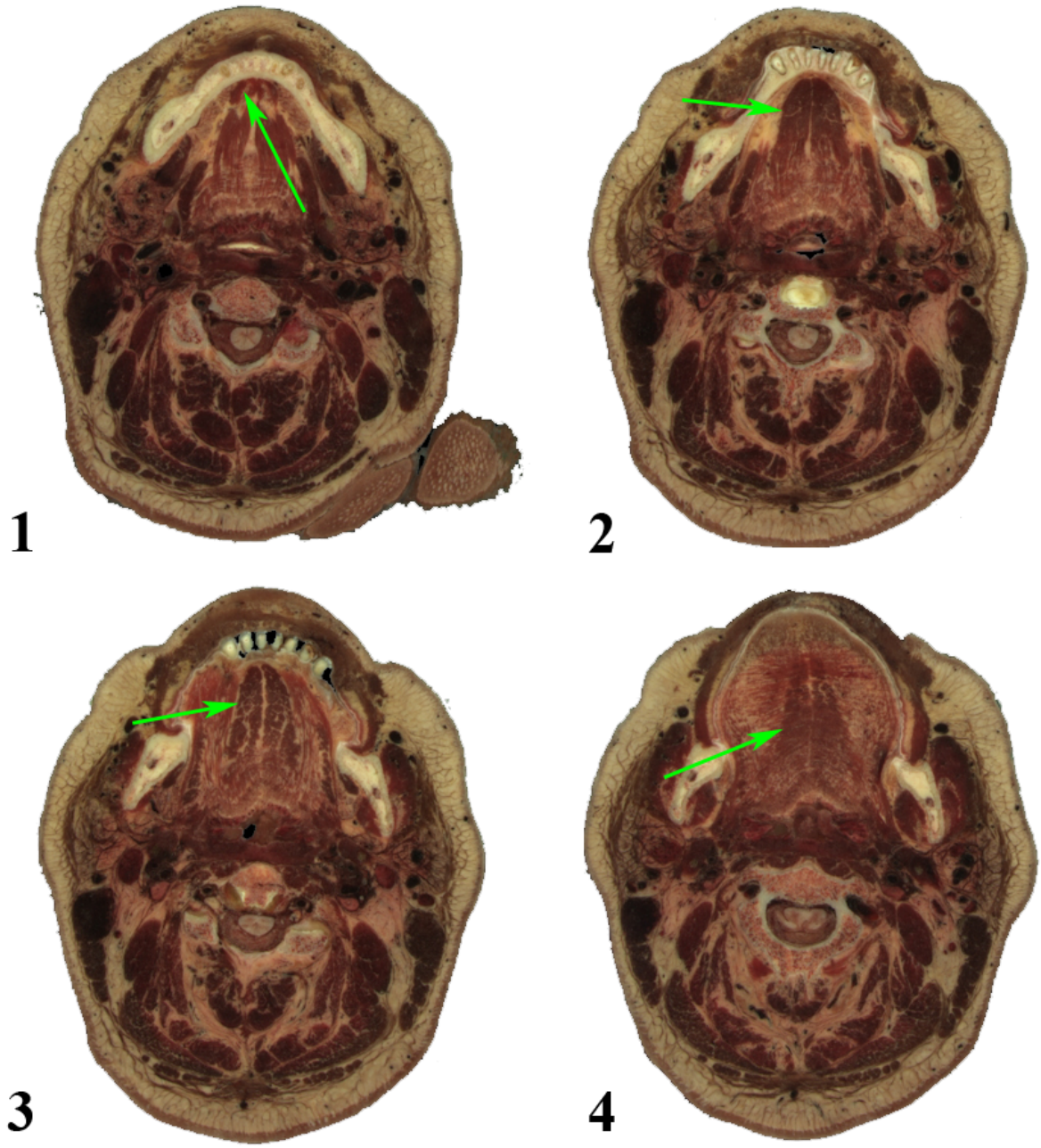


Figure 7.8: Axial VHF view of the genioglossus. The genioglossus is highlighted by green arrows. Image 1 shows the origin of the GG at the symphisis menti. In image 2, the GG appears as a uniform mass of muscle as it fans outwards. In image 3, fibrofatty infiltration of the GG is evident as it moves superiorly and posteriorly in the oral cavity. In image 4, the GG is barely distinguishable from the intrinsic muscles of the tongue. The fibres of the transversus muscle stand out as they run from left to right.

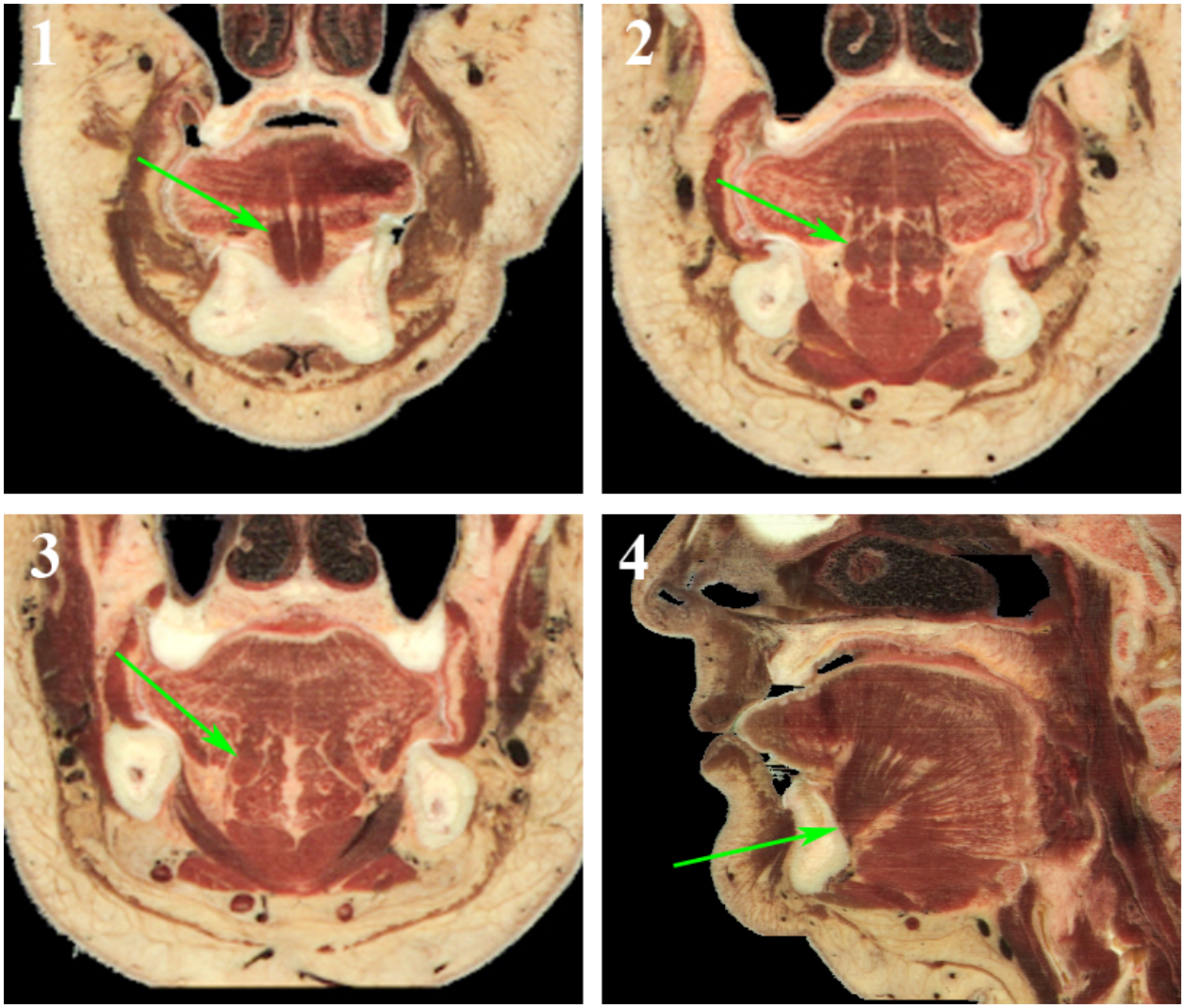


Figure 7.9: Coronal and sagittal VHF views of the genioglossus. The genioglossus is highlighted by green arrows. Image 1 shows the origin of the GG at the symphisis menti. In image 2, the GG is darker in color and superior to the geniohyoid. In image 3, fibrofatty infiltration of the GG is evident as it moves superiorly and posteriorly in the oral cavity. Image 4 is a sagittal view showing the fan-like structure for which the GG is known. The green arrow indicates the origin of the "fan" at the symphysis menti.

7.4.1.4 The palatoglossus

The small, thin palatoglossus is difficult to identify. In the VHF, it is best seen in the axial plane. It can be identified at its origin on the anterior surface of the hard palate. The muscle then thickens as it descends inferolaterally, passing anterior to the palatine tonsil to reach the posterolateral tongue. Here, palatoglossus is seen to merge with both the intrinsic muscles (particularly transversus and superior longitudinalis) as well as styloglossus.

Like the other three extrinsic muscles described, at the origin and through much of its course, palatoglossus is uniformly deep red in hue with minimal fibrofatty infiltration. Upon reaching the lateral tongue, its integration is abrupt. This, coupled with the small size of the muscle, means that no further comment on compositional changes can be made.

Axial views of the palatoglossus are described in Figure 7.10.

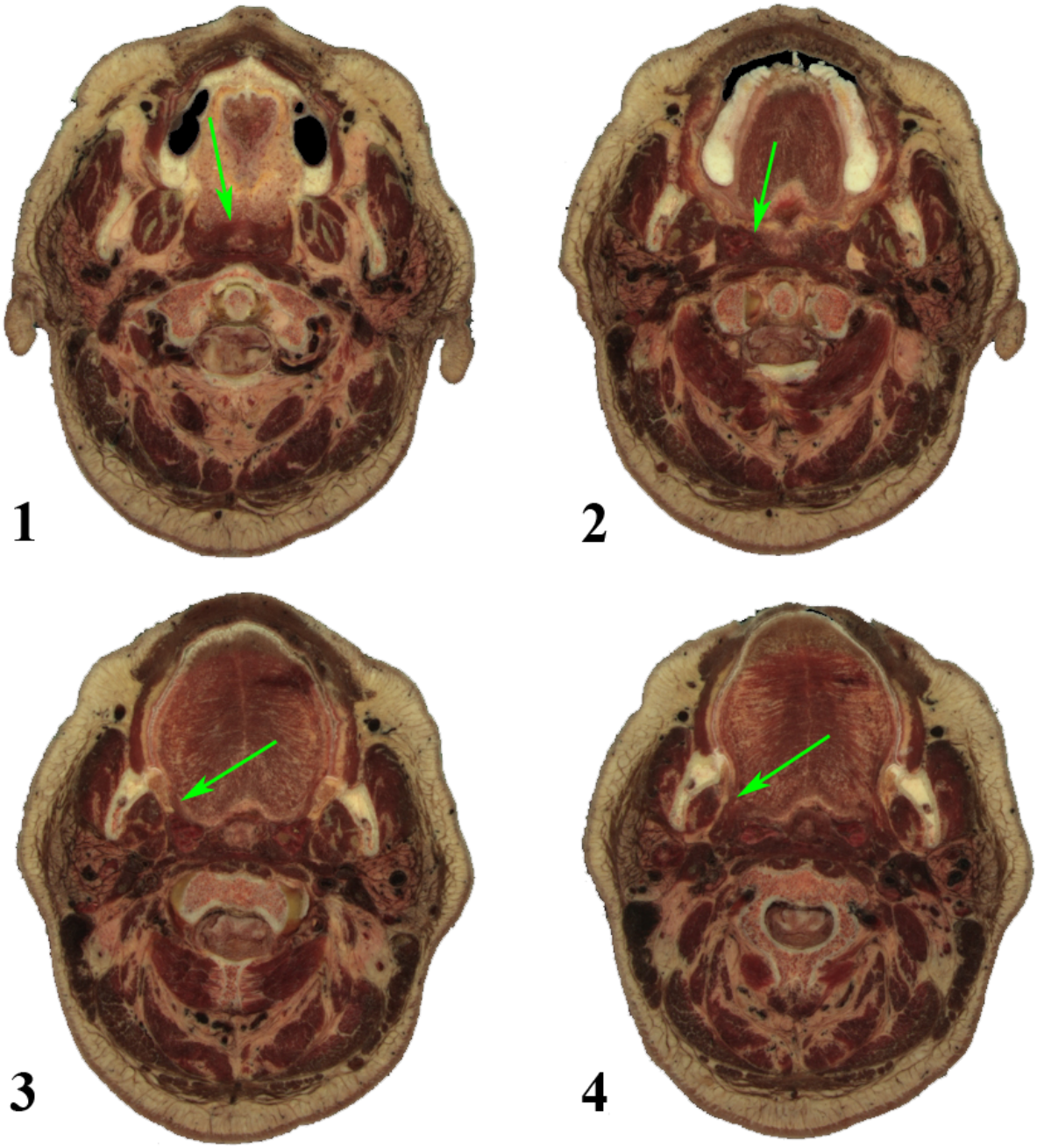


Figure 7.10: Axial VHF view of the palatoglossus. The palatoglossus is highlighted by green arrows. The palatoglossus is a small, thin muscle that is difficult to view in any plane. Image 1 shows the PG at its origin at the posterior margin of the hard palate. Image 2 and 3 show the PG as it descends inferolaterally, passing anterior to the palatine tonsil. Image 4 shows the PG as merges with the posterolateral tongue.

7.4.1.5 The intrinsic muscles of the tongue

The intrinsic tongue muscles of the VHF are distinctive in appearance. The intrinsic muscles appear as a lighter hue of red compared to the darker extrinsic muscles. The intrinsic muscles also appear less homogeneous in composition due to heavily fatty infiltration. This fatty infiltration allows the observer to distinguish, to some extent, muscle fiber directionality. The transversus appears particularly dominant and is easily seen in both axial and coronal views.

7.4.2 Distance of extrinsic muscles from the surface mucosa

The previous section described the anatomy of the extrinsic lingual musculature as seen on the VHF. Notable in these findings was the proximity of the styloglossus and hyoglossus to the surface of the lateral tongue. The following section describes the result of measurements taken to qualitatively evaluate this relationship. Table 7.1 summarizes the results.

Table 7.1: Measurement data for the distance of the lateral extrinsic muscles from the lingual surface. The mean measurement value is given, accompanied by the range, for the left and right lateral extrinsic muscles of the tongue categorized by viewing plane. The "combined" category combines the results of the axial and coronal measurements. All measurement units are in millimeters (mm).

	Styloglossus	*Hyoglossus*
Axial		
Right	*1.33 mm [0.33 - 1.48]*	*2.93 mm [1.48 - 4.96]*
Left	*2.91 mm [0.66 - 7.68]*	*4.33 mm [1.68 - 8.71]*
Coronal		
Right	*0.84 mm [0.33 - 1.68]*	*1.85 mm [0.74 - 3.69]*
Left	*2.30 mm [0.66 - 4.87]*	*1.63 mm [0.99 - 3.04]*
Combined		
Right	*1.08 mm [0.33 -3.69]*	*2.39 mm [0.74 - 4.96]*
Left	*3.18 mm [0.66 - 7.68]*	*3.39 mm [0.99 - 8.71]*
Both	*1.84 mm [0.33 - 7.68]*	*2.69 mm [0.74 - 8.71]*

7.4.2.1 The styloglossus

In the axial plane, the distance from the mucosal surface to the superficial border of the styloglossus was measured in seven consecutive slices. The same series of slices were used for measurements

concerning both the left and right styloglossi. On the right it was found that the mean of the distances measured was 1.33 mm with a range of 0.33 mm to 3.69 mm. On the left, the mean distance was 2.91 mm with a range of 0.66 mm to 7.68 mm.

The maximum distance from the mucosal surface to the styloglossus on the left was increased by a superior extension of the sublingual gland, highlighted in Figure 7.11. This glandular mass was located deep to the mucosa but superficial to both lateral muscles, resulting in abnormally large maximum distances. This is illustrated in Table 7.1.

Similar results were found in the coronal plane. 16 consecutive slices were used to measure the distance of the styloglossus. Again, the same slices were used on the left and right. On the right the mean of the distances measured was 0.84 mm with a range of 0.33 mm to 1.68 mm. On the left, the mean distance was 2.30 mm with a range of 0.66 mm to 4.87 mm.

Combining the results from both planes, the styloglossus was a minimum of 0.33 mm from the mucosal surface. The maximum and mean distances measured were 7.68 mm and 1.48 mm respectively. If the left sided measurements are excluded due to the presence of the fat-containing mass, the maximum and mean distances become 1.33 mm and 1.08 mm respectively.

7.4.2.2 The hyoglossus

In the axial plane, the distance from the mucosal surface to the most superficial fibres of the hyoglossus was measured in 8 consecutive slices on the right and 7 consecutive slices on the left. On the right the mean distances measured was 4.96 mm with a range of 1.48 mm to 4.96 mm. On the left the mean distance was 4.33 mm with a range of 1.68 mm to 8.71 mm.

Similar results were found in the coronal plane. 12 consecutive slices were used to measure the distance of the styloglossus. Again, the same slices were used on the left and right. On the right the mean of the distances measured was 1.85 mm with a range of 0.74 mm to 3.69 mm. On the left the mean distance was 1.64 mm with a range of 0.99 mm to 3.04 mm.

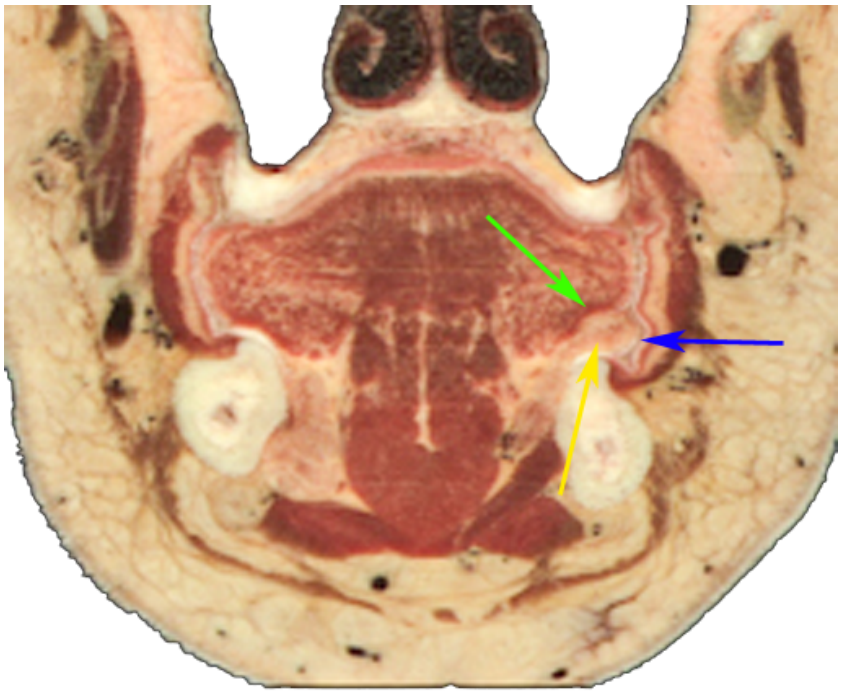


Figure 7.11: Fatty mass on lateral tongue. The mass overlies the styloglossus and hyoglossus, increasing the measured distance from the surface. The mass is thought to be a superior extension of the sublingual gland. Green arrow: styloglossus; yellow arrow: fatty mass; blue arrow: lingual sulcus.

Combining the results from both planes, the hyoglossus was a minimum of 0.74 mm from the mucosal surface. The maximum and mean distances measured were 8.71 mm and 2.15 mm respectively. If the left sided measurements are excluded due to the presence of the fat-containing mass, the maximum and mean distances were 4.96 mm and 2.39 mm respectively.

7.5 Discussion

This section brings the experimental work together to discuss the implications of the findings. Shortcomings of the work are also identified.

7.5.1 The lingual anatomy of the VHF

The lingual anatomy of the VHF does not depart appreciably from the classical descriptions of Abd-El-Malek and Miyawaki [223-224]. This is the same description that has been adopted by anatomy texts [2].

In fact, the VHF displays remarkable symmetry in both position and dimension of the extrinsic and intrinsic muscles. The most notable exception lies in the morphology of the sublingual glands. The left sublingual gland extends further superiorly and posteriorly than the right. In its superior extent the left gland is found in the lateral tongue, superficial to the styloglossus, thus increasing the apparent distance from the surface.

Notably, the styloglossus and hyoglossus muscles appear to be distinct, separated by a thin fascial plane. This is contrary to the view of the Golding et al. that interdigitation of styloglossus and hyoglossus in the anterior tongue forms a combined lateral extrinsic muscle group [228]. While anatomically separate, it is likely that the close juxtaposition of the styloglossus and hyoglossus makes them appear as a single unit on MRI. The combination of the two muscles into a single lateral group is a useful paradigm for interpreting diagnostic imaging studies.

Diffusion tensor imaging (DTI) tractography, shown in Figure 7.12, is able to distinguish hyoglossus and styloglossus. A 2007 study by Gaige et al. demonstrated that DTI is able to segment the

styloglossus and hyoglossus based on muscle fiber orientation [229]. Gaige also commented on the extent of the styloglossus stating:

> *The green longitudinal tracts confirm the presence of a thick styloglossus (sg) insertion into the tongue body that merges with the inferior longitudinalis, coursing along the lateral edges and wrapping around the anterior tip of the tongue.*

This statement highlights an important point. The styloglossus continues beyond the interface between the hyoglossus and tongue body until it reaches the very tip. On the VHF, the styloglossus could not be distinguished all the way to the tip. This may be an anatomical variation of the VHF. However, it may also be that the anterior extent of the styloglossus is beyond the resolution of the VHF images and only visible on DTI tractography.

Because the description of the VHF lingual myoarchitecture correlates with classical descriptions, it is suitable for morphologic study of the relative distance of the lateral extrinsic muscles from the lingual surface. Further, this also confirms that the VHF was a suitable choice for construction of an anatomic atlas, completed previously in our Department. Considering that the VHF used high resolution photos, 0.33 mm slices and rigorous conditions, the VHF represents a more accurate medium for anatomical study. However, we must also consider that the classical anatomic studies were carried out by highly experienced specialists in oral anatomy, are based on multiple specimens, and were done under direct vision as opposed to digital imaging.

7.5.2 The position of the lateral extrinsic muscles: implications

The results of the distance of the styloglossus and hyoglossus have several implications for OCSCC TNM staging. First of all, the lateral extrinsic muscles of the tongue, particularly the styloglossus, are very close to the surface. Based on measurements, it could be argued that the styloglossus is within a millimeter of the lingual surface throughout much of its course. Further, DTI tractography demonstrates that the superficial course of the styloglossus may extend from base to apex [229].

Figure 7.12: DTI tractography of the lingual myoarchetecture. The left image is taken from Grey's Anatomy and shows the extrinsic muscles of the tongue. This is compared to DTI tractography on the right. Green lines represent muscle fibres running anterior-posterior, red transverse and blue superior-inferior. Fibres running oblique to the three orthogonal planes are represented by combinations of these colours. gg – genioglossus; gh – geniohyoid; hg – hyoglossus; il – inferior longitudinalis; pg – palatoglossus; sg – styloglossus; sl – superior longitudinalis; v – verticalis. Adapted from Gaige et al. Source: [229]

7.5.2.1 The interpretation of TNM staging criteria

If we interpret the TNM criteria as referring to all of the extrinsic muscles, it is likely that almost all lateral tongue SCCs would qualify as T4 tumours due to lateral muscle invasion. If we also consider that the majority of oral cancers are lingual, and the lateral tongue is the most common site for SCC, a strict interpretation of TNM criteria would lead to upstaging of a considerable number of tumours.

In the original publication of TNM staging in 1977, invasion of the tongue "root" is identified as a T4 criterion [230]. This definition was altered in the second edition, published in 1983, in which invasion of the "base" of the tongue was identified [231].

This definition was further altered in the third edition in 1988 to mean the “deep [extrinsic muscles]”, marking the first appearance of specific reference to the extrinsic muscles [232]. In many ways, the 1988 T4 criterion is ambiguous. No specific muscles are mentioned and interpretation is left up to the surgeon or pathologist. "Deep" may be interpreted to mean the genioglossi only, or perhaps the genioglossi and hyoglossi.

The "deep [extrinsic]" definition persisted in the fourth edition (1992) and fifth edition (1997) [233-234]. The definition of "deep [extrinsic]" was expanded in the sixth edition (2002) to specifically refer to the "genioglossus, hyoglossus, palatoglossus, and styloglossus" [190]. This definition persists in the current seventh edition, recently published in December of 2009 [69].

While this clarification is helpful, it introduces further problems with interpretation. The extrinsic muscles cannot be globally referred to as "deep". Each muscle occupies a spectrum from "deep" to superficial. For example, the genioglossus might be referred to as deep within the tongue bulk, but in the floor its anterior fibres are relatively superficial.

7.5.3 Extrinsic muscle invasion: clinical value

Despite the inclusion of extrinsic muscle invasion in TNM criteria since their inception, it is difficult to discover why exactly they are a poor prognostic indicator. The only information available is that the

extrinsic muscles act as a route of invasion to the floor and other deep structures of the head and neck. There is little literature support for this mechanism. Like bone invasion, it may be that the "invasion conduit" theory arose from consensus agreement when the TNM criteria were originally created and, also like bone invasion, over the years has become entrenched [106].

If we assume this to be true, invasion of the extrinsic muscles should result in predictable patterns of tumour extension. Lateral tongue tumours should follow the course of the styloglossus anteriorly and posteriorly. Posterior extension along the styloglossus would extend to structures near the skull base. Following the hyoglossus, the tumour would track inferiorly to invade the floor. Lingual carcinomas that invade deeply to reach the genioglossus could follow the muscles anteriorly to reach the symphysis menti, posteriorly to reach the tongue base or inferiorly to reach the floor.

For small tumours invading the styloglossus it seems unlikely that the T4a designation is justified, even if the "conduit" theory is true. However, such concern may be justified in the tongue base where the styloglossus ascends into the surgically challenging structures of the posterior pharynx. Certainly, the T4a designation is inappropriate for superficial tumours of the anterior two-thirds of the tongue. Similarly, tumours invading only 1-2 mm may invade the hyoglossus, making T4 designation inappropriate, despite the potential to spread to the floor.

7.5.4 Shortcomings

7.5.4.1 VHF – anatomy based on a single subject

The VHF represents a single person and is far from representative of the entire population. However, this is also a problem with classical descriptions. Miyawaki bases his observations on three human specimens [224] while Abd-El-Malek bases his conclusions on a small, but unspecified, number [223].

The fact that there is agreement between the three descriptions lends validity to the observations. Undoubtedly there are a large number of anatomic variations in lingual anatomy. A comprehensive summary of all anatomic variants would require a large number of specimens.

7.5.4.2 VHF – preservation alters tissue dimensions

The VHF was first embedded, unfrozen, in quickly hardening foam to allow for CT and MRI scanning. The VHF was then frozen to -70°C to allow for sectioning into four blocks with three saw cuts. After sectioning, the hardened foam was mostly removed and replaced by a 3% gelatin solution which was frozen and maintained at -70° for the duration of the 12 months it took to section the VHF [235].

It is unknown how this process affected the VHF. At no point was the VHF preserved in formalin, which can lead to significant soft tissue shrinkage. It is still likely that some tissue shrinkage occurred, simply from desiccation. This is of concern for measurements of the distance of the extrinsic muscles. Lack of perfusion in the post-mortem period would also lead to shrinkage. The result of this hypothesized shrinkage is that measured distances may underestimate *in vivo* dimensions.

7.5.4.3 The arbitrary choice of measurement locations

The muscle distance measurements must also be interpreted with a measure of caution. While an attempt was made to take measurements from a variety of positions on the lateral tongue, the choice of measurement position was partly arbitrary. Therefore, while the mean measurement is difficult to interpret, it does give the reader some idea of the difference between the distance of the styloglossus and the hyoglossus. This arbitrary choice of measurement location also leads to the large ranges seen in Table 7.1. In some cases the range of measurements was up to 8 mm.

On the left side the maximum and average distance were inflated by the superior position of the left sublingual gland. However, the critical measurement is the minimum distance. The minimum is important for drawing attention to the fact that the styloglossus and hyoglossus, in some places, lie within 1 mm of the lingual surface.

7.6 Conclusions

In this chapter we have seen that the Visible Human Female dataset is a highly detailed source of anatomic information describing the oral cavity. Anatomic descriptions derived from the VHF are similar to those of classical anatomists and, in many cases, provide added detail.

From VHF-based measurements of the oral cavity anatomy, it is obvious that the styloglossus and hyoglossus muscles are very close to the surface of the lateral tongue. Even very superficial tumours of the lateral tongue are likely to invade the extrinsic muscles. Such tumours are unlikely to warrant a T4a designation and the aggressive treatment that such a designation entails. While the AJCC specifies that "deep" extrinsic muscle invasion is required for T4a designation, this description is both ambiguous and inadequate.

The lateral tongue is a common site for oral carcinoma and, therefore, tumours in this area require careful consideration and accurate staging. The following Chapter is an extension of work in this Chapter, and explores the concept of lateral extrinsic muscle invasion further through the use of an anatomic atlas. The atlas was developed in house from the VHF and shows promise for illustrating extrinsic muscle invasion through fusion with MRI of OCSCC.

8 VHF tongue atlas: clinical application

8.1 Introduction

The previous chapter established that the lateral extrinsic muscles lie very close to the lingual surface. Theoretically, this means that small tumours of the lateral tongue have the potential to invade the styloglossus and hyoglossus very early in their natural history. Strictly using the AJCC definition of "extrinsic muscle" included as part of the oral cavity T4a classification, many tumours would be inappropriately up-staged.

It is difficult to assess the invasion of the lateral extrinsic muscles of the tongue on MRI. These muscles are often poorly seen, especially on T2-weighted images used to highlight invasive carcinoma. A system to evaluate lateral muscle invasion during pre-surgical staging requires some sort of visualization method. Previous work in this Department established an anatomic atlas of the extrinsic muscles of the tongue derived from the Visible Human Female. The VHF has already been established as a viable representative of lingual anatomy, therefore an atlas derived from the VHF is likely to be a valuable resource.

Theoretically, non-rigid fusion of the extrinsic muscle atlas with MRI imaging of lingual SCC should provide a rough, yet practical, means by which to visually assess invasion of the extrinsic muscles of the tongue. This chapter aims not only to establish methodology through which tongue-atlas fusion may be achieved in a clinical setting, but also to assess if lateral extrinsic muscle invasion is of any practical value when staging lingual carcinoma. Implications to TNM staging criteria are also assessed.

8.1.1 Chapter goals

The goals of this chapter are:

1. To give an outline of previous work in the Department that led to the creation of a VHF-based anatomic atlas of the extrinsic muscles of the tongue.

2. To develop a procedure with which the extrinsic muscle atlas can be fused with coronal MRI of the oral cavity.

3. To use fusion of the tongue atlas with MRI series of lingual SCC.

4. To use these MRI-atlas fusions to establish when invasion of the extrinsic muscles of the tongue has taken place

5. To evaluate the prognostic significance of lateral extrinsic muscle invasion and genioglossus invasion as determined by the MRI-atlas fusions.

6. To identify shortcomings of the current TNM T4a criteria based on invasion of the "deep" extrinsic muscles and to suggest alternate wording if appropriate.

8.1.2 Organization of sections

This chapter begins with a brief outline of techniques for identifying and imaging the muscular anatomy of the tongue. Previous work in the Department building a VHF-based anatomic atlas of the extrinsic muscles of the tongue is discussed.

This VHF tongue atlas is adapted for co-registration with MRI of the oral cavity. Subsequently, co-registration is carried out using a retrospective series of 87 patients with SCC of the oral tongue. TNM staging is re-evaluated following evaluation of lateral extrinsic muscle invasion.

The implications of potential staging problems created by a strict definition of "extrinsic muscle" as offered by the AJCC are discussed and potential modifications to this definition recommended. The chapter ends with a brief summary and conclusions.

8.2 Background

The anatomy of the extrinsic muscles of the tongue, as well as their role in TNM staging, were detailed in the background of Chapter 7. This section focused on the identification of extrinsic muscles on imaging, surgery and pathology, previous work on the VHF-based atlas and fusion with MRI of the oral cavity.

8.2.1 Identification of the extrinsic muscles

Given the role of extrinsic muscles in T4a classification, reliable methods are required to identify

invasion. The extrinsic muscles can be partially identified on pre-surgical imaging. Superb soft tissue resolution makes MRI the modality of choice.

8.2.1.1 Magnetic resonance imaging

Historically, there has been greater interest in MRI of the tongue than other oral cavity subsites. Early work by Unger was able to identify the genioglossus and geniohyoid muscles [62]. Lufkin followed a year later in 1986 with improved technology, and was able to demonstrate the superior longitudinal, genioglossus and transverse muscles with increased spatial resolution [63]. MRI of the oral tongue is reviewed in Chapter 2. Examples are shown in Figure 8.1 and Figure 8.2.

In work completed in this Department, the visibility of the lateral extrinsic muscles was evaluated using the T1-weighted SE images of 125 patients with suspected salivary gland pathology [228]. The hyoglossus was seen in 88% of coronal series and 95% of axial series. Similarly, the styloglossus was seen in 88% of coronal and 82% of axial series. The palatoglossus muscle was not visualized. While the styloglossus was frequently seen, detection of its presence was limited to the segment found between its insertion at the styloid process and the tongue. Its presence in the lateral tongue, as suggested by classical descriptions, was not documented. Instead, the author evaluated the existence of a lateral muscle group. Such a group was postulated by Golding et al. and was thought to consist of the interdigitation of fibres of the hyoglossus and styloglossus [228]. This complex was visible in only 30% of coronal and 24% of axial series.

8.2.1.2 Computed tomography

According to Larsson et al., the hyoglossus, genioglossus and styloglossus can all be identified on CT. The palatoglossus is not mentioned, though it is likely that the small size of this muscle makes identification difficult [236]. A case series of 35 patients by Muraki et al. examined the ability of CT to identify structures in the floor and oropharyngeal tongue. The genioglossus-geniohyoid pair was identified in all 35 cases, while the styloglossus-hyoglossus pair was identified in 31 cases [237].

The intrinsic muscles of the tongue are identified as "a midline, rounded ball of muscle without

fascial planes or regular direction of muscle bundles" [236]. According to Muraki et al., the intrinsic muscles were visible in all 35 patients and were generally 10-20 Hounsfield Units (HU) lower in density relative to the extrinsic muscles [237]. An example is shown in Figure 8.3.

8.2.1.3 Surgery

The primary concern of the surgeon is to achieve resection margins that are clear of tumour. While resection of the extrinsic muscles of the tongue leads to functional impairment in speech and swallowing, this concern is secondary to curative intent. Nevertheless, the surgeon may be able to identify frank invasion of extrinsic muscles at the time of surgery, altering the TNM staging of the primary tumour. However, it is preferable to have this information prior to surgery to aid surgical planning and to anticipate functional defects.

8.2.1.4 Pathology

The work of Miyawaki, shown in Figure 8.4, was carried out using whole human tongues sliced into coronal, sagittal or axial slices 2-3 mm thick. The tongues were prepared first by formalin fixation, then ethanol dehydration and finally xyrol, to render the muscle fibres transparent [224].

Takemoto took wedge shaped sections of the tongue and prepared them in a fashion similar to Miyawaki. He used his findings to construct a three-dimensional model that supports classical anatomic descriptions [238].

The pathologic description of lingual myoarchitecture is possible on specimens of whole tongues. However, problems are more pronounced on standard pathology slides prepared from resected surgical specimens. Thin slicing, difficulty in spatial orientation and distortion caused by the invasive tumour make identification of extrinsic muscles, much less the evaluation of extrinsic muscle invasion, very difficult (personal communication, Dr. K. Shah, Oxford). While pathological evaluation is a cornerstone of TNM staging, little information is offered about extrinsic muscle invasion.

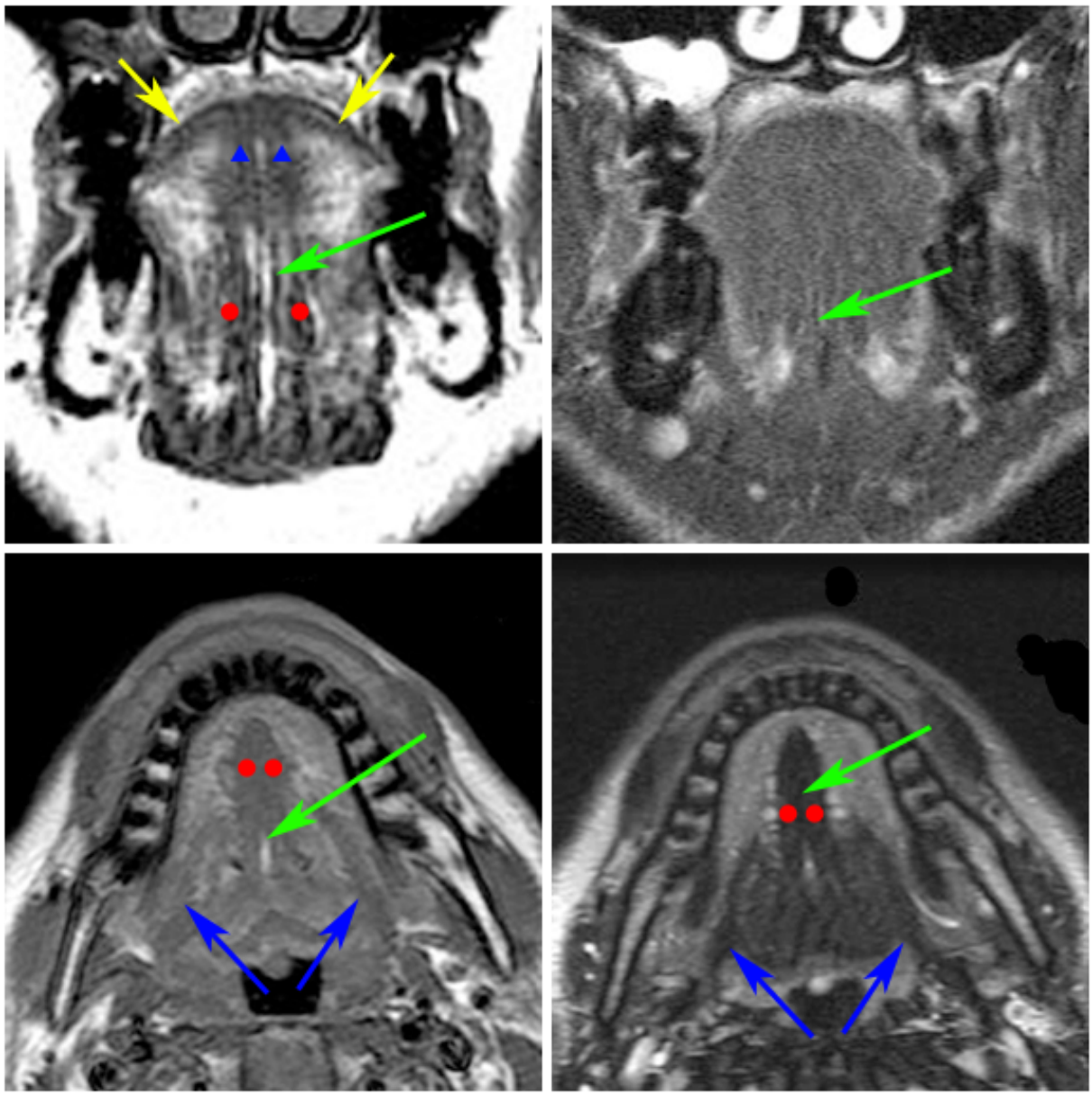


Figure 8.1: Coronal and axial MRI views of the normal tongue. Top left: T1-weighted fast-spin echo view of the oral cavity, taken at mid orbital level, demonstrates several aspects of the oral anatomy. The midline lingual septum is visible as a vertical line of high signal intensity (green arrows). On either side of the lingual septum the genioglossi are visible as wedge-shaped areas of low signal intensity (red circles). The transversus (blue triangle) and superior longitudinal muscles (yellow arrows) are also seen. Top right: T2-weighted, fat-suppressed FSE image shows the tongue as a homogeneous area of low to moderate signal intensity. Bottom left: T1-weighted fast-spin echo image taken at the level of the mental spine. Several aspects of lingual anatomy already mentioned are visible. Notable is the hyoglossus as an elongated region of low signal intensity (blue arrows). On the right, a T2-weighted image at the same level.

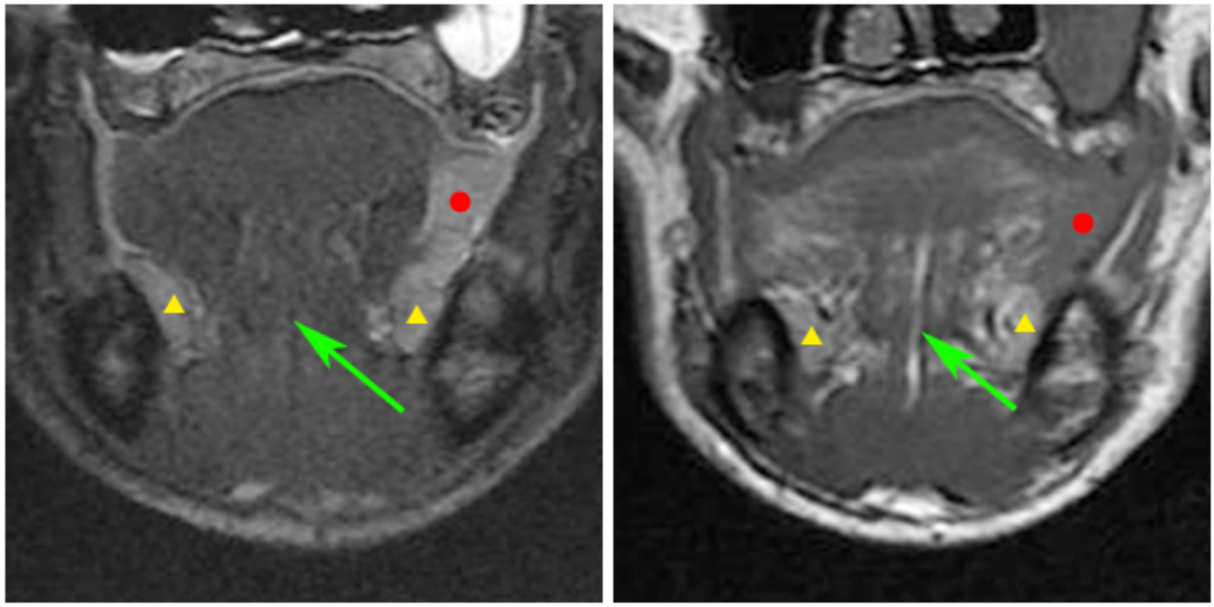


Figure 8.2: T2-weighted and T1-weighted views of carcinoma of the tongue. Both the left and right images show a patient with a T2 squamous cell carcinoma of the left lateral oral tongue (red circles). The left shows a T2-weighted coronal image in which the tumour is clearly demonstrated as bulbous region of high signal intensity. On the right, the tumour is less conspicuous as a region of low signal intensity. However, the T1-weighted image shows the internal anatomy of the tongue with superior detail. Green arrow: lingual septum; yellow triangle: sublingual gland.

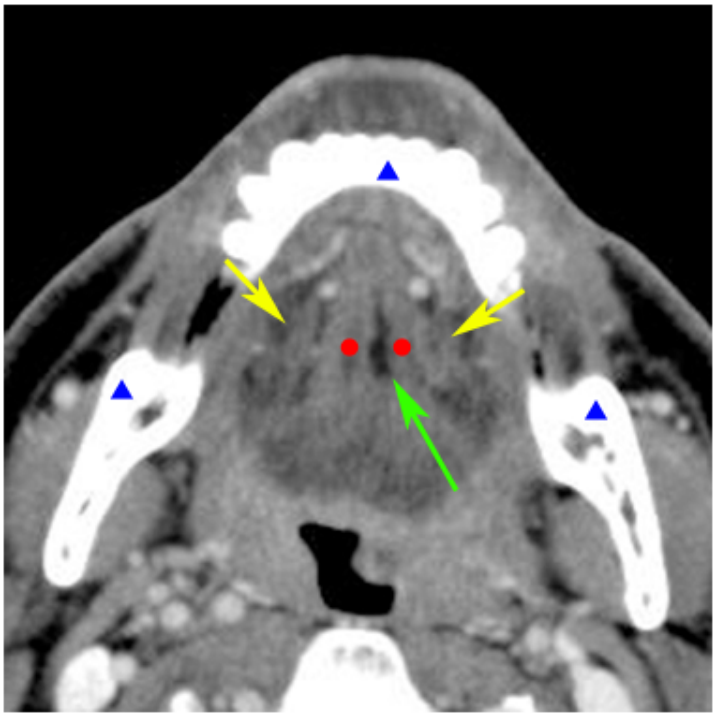


Figure 8.3: CT image of tongue. This contrast-enhanced CT image uses a soft tissue window specified by the eFilm software used for viewing. The cortical bone of the mandible is well defined (blue triangles), as are vascular structures, as high density regions. Soft tissues appear as a limited range of shades of grey. The genioglossi (red circles) and hyoglossi (yellow arrows) are seen, as is the lingual septum (green arrow).

Figure 8.4: Original drawings of Miyawaki. Miyawaki's drawings were pathological analysis of cadaver tongues. These two sketches, labeled S:1 and S:2 representing sagittal slices 1 and 2, show the muscular anatomy of the tongue slightly to the right of the midline. Superb detail of the directionality of the extrinsic and intrinsic muscles fibres is demonstrated. GG: genioglossus; T: transversus, IL: inferior longitudinalis; SG: styloglossus; HG: hyoglossus; V: verticalis; SL: superior longitudinalis. Source: [224]

8.2.2 Previous work with the VHF atlas

Previous work in this Department used the VHF to construct an atlas of the extrinsic muscles. The work also made some progress towards automatic registration of the atlas with clinical MRI scans. It was successful in a proof of principle validation using bovine tongue, as well as ideally prepared human volunteers [1]. The next phase of the work is the transition to clinical application. Examples of atlas slices are shown in Figure 8.5.

8.3 Methods

In the following sections, the implementation of the VHF-based anatomic atlas of the tongue is explored as a device for assessing tumour invasion of the extrinsic muscles.

8.3.1 Identification of case series

Patients with carcinoma of the oral tongue were identified from the existing database outlined in Chapter 3. Details of the patient population are available in Chapter 4. Patients were excluded if their MRI images were of insufficient quality to fully define the contour of the tongue.

MRI series were viewed in eFilm. All available pulse-sequences were used to evaluate image quality for susceptibility and motion artefact. Comparison between pulse-sequences was used to identify confounding features, such as tumour bordering or invading sublingual glands. Radiology reports, when available, were used to aid image interpretation. Images were viewed by a single viewer, the author, using hardware identified in the general methodology in Chapter 3.

After global review of the available images, tumour-containing slices of T2-weighted series were identified and exported as JPEG files. Where there was doubt as to the extent of tumour growth, images were reviewed by a consultant radiologist with experience in head and neck imaging (SJG). If T2-weighted series were unavailable or of poor quality, alternate pulse-sequences were not used and the case was excluded. Because the atlas is only available in the coronal orientation, only coronal studies were used for analysis.

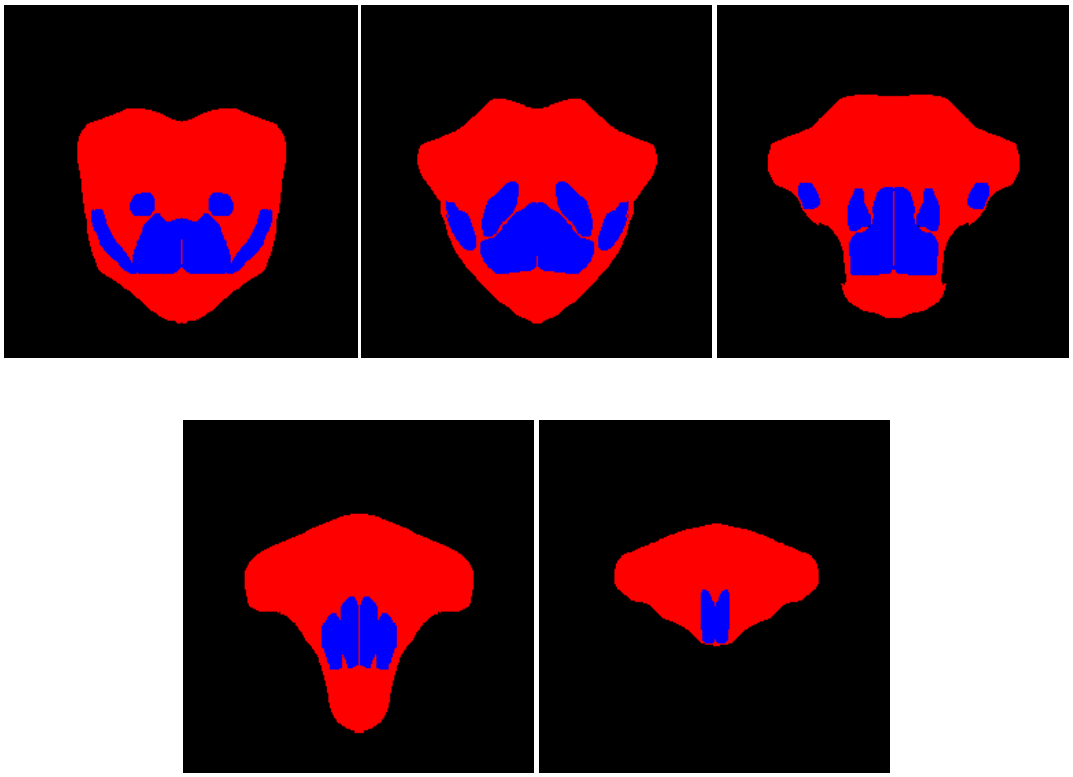


Figure 8.5: The VHF tongue atlas. A series of images taken from the VHF-derived tongue atlas, developed by previous work in our Department, are shown. The extrinsic muscles are shown in blue while the remainder of the tongue is shown in red. These images correspond to, beginning in the tongue base, slices 1, 27, 55, 92 and 116 of the atlas [1].

8.3.2 VHF atlas pre-processing

There were several problems with the VHF atlas. When the VHF died, rigor mortis was allowed to set in with the mouth open and the tongue protruding beyond the lips. Also, the VHF, with the exception of teeth 1-3 on the lower right and 1-4 on the lower left, is edentulous. Because the mandible is depressed and the mouth open, the tongue protrudes laterally into the resulting space between the upper and lower alveoli. The majority of patients undergoing MRI do so with the mouth completely closed and are not edentulous. Thus, the scanned tongue differs in shape from the atlas tongue.

Because the VHP tongue is protruding, it is elongated in the AP axis relative to the imaged tongue. This elongation appears more pronounced in the anterior one-third of the tongue and less pronounced in the base. Similarly, the depressed, edentulous mandible results in a relative increase in the left-right diameter of the tongue manifest as a "winging" of the lingual volume. These shortcomings are shown in Figure 8.6.

The tongue can be thought of as a muscular hydrostat. Therefore, assuming that the volume of the VHP and imaged tongues are constant, there must be compensatory changes in shape when confined to the closed-mouth oral cavity. In the presence of complete dentition, closing the mouth results in the loss of "winging" and the reduction in the AP diameter of the tongue. This is compensated for by a gain in the cranial-caudal diameter, facilitated by the tongue fully occupying the space between the dorsum and the palates, as well as a fall in the floor of the oral cavity. Further, in the closed mouth the tongue protrudes into the oropharynx. The issue of oddly-shaped tongue was addressed by excluding any atlas slices anterior to the labial margin of the mandibular incisors. 60 slices were removed, leaving a total of 130 slices in the revised atlas.

The choice of atlas image is somewhat subjective. The Z-axis resolution of the atlas (0.33 mm) exceed that of the MRI series (typically 3-6 mm). Therefore, anywhere between 10 and 20 atlas slices could represent a single MRI slice. The number of slices was reduced by two-thirds by only including every third slice in the final version. Thus, the final version of the atlas contained 44 distinct slices.

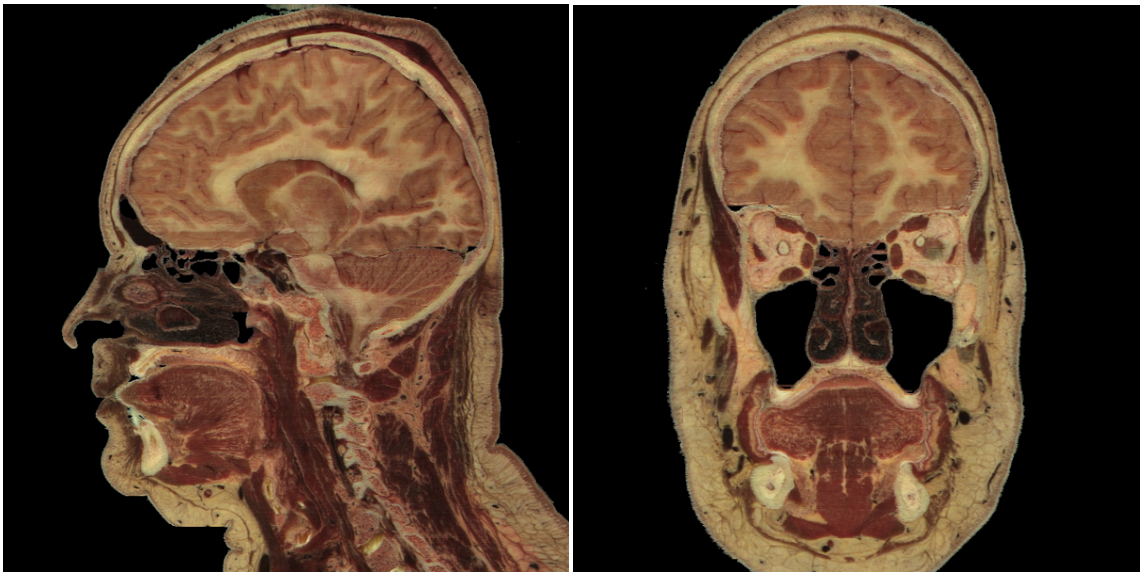


Figure 8.6: The anatomic positioning of the VHF tongue. On the left, a sagittal view of the VHF demonstrates the protrusion of the tongue apex beyond the limits of the anterior dentition. On the right, a coronal view shows the "winging" of the tongue created by the loss of the posterior dentition in the VHF female. Both of these features are uncommonly found in MRI of the head and neck [226].

8.3.3 Sample characteristics

A total of 97 patients with SCC of the oral tongue were identified from the main database. 9 patients were excluded because the T2-weighted coronal scans were of poor quality due to motion or susceptibility artefact. 1 case was excluded because the T2-weighted coronal MRI series was corrupt.

87 patients remained for statistical analysis. 16 (18.4%) of these cases had superficial tumours that were not visible on MRI and therefore were designated as negative for extrinsic muscle involvement. The 71 (21.6%) patients remaining comprised a total of 292 tumour-containing MRI slices to which VHF-atlas fusion was carried out.

8.3.4 Selection of the appropriate atlas slice

The appropriate slice number was chosen by multiplying the total number of tongue-containing atlas slices (44) by the ratio of the distance of the tumour containing MRI slice from the posterior limit of the tongue to the anteroposterior diameter of the tongue. The result was then rounded to the nearest whole number slice. The equation is given below:

Equation 8-1

$$\frac{d_S}{d_{AP}} \times n_{atlas} = s_{fusion}$$

where d_S is the distance of the tumour containing slice from the posterior tongue margin, d_{AP} is the AP diameter of the tongue, n_{atlas} is the number of slices comprising the atlas and s_{fusion} is the slice number to be used for tongue-atlas fusion.

8.3.5 MRI - atlas fusion

The MRI-derived JPEG and atlas slices were imported into FantaMorph (v. 4.1, Abrosoft), a commercially available image fusion software. FantaMorph uses a non-rigid image co-registration algorithm to warp a source image onto a recipient image. In this case, the source image was the atlas slice and the recipient image was the tumour-containing MRI slice.

The FantaMorph software used a mesh-based algorithm for which the user must manually assign a series of corresponding control points in both the source and recipient image. The exact algorithm is proprietary information but it appears to be based on simple mesh deformation. The end result is the deformation of the VHF atlas control points to the corresponding MRI image control points [239]. An example is given in Figure 8.7.

A maximum of 20 control points were used to perform the non-rigid deformation. The majority of these control points were placed on the surface of the tongue. The first four control points were placed at the cardinal directions, defining the superior, inferior and lateral limits of the tongue contour. The remaining control points were evenly spaced throughout the remainder of the tongue contour. Where the midline of the tongue was visible, control points were added to compensate for deformation of the midline by tumour mass effect. The deformed atlas slice was then overlaid onto the MRI image to delineate the theoretical position of the extrinsic muscles on MRI. The resulting composite image was exported as a JPEG file and stored for reference. The presence of tumour invasion of the extrinsic muscles was assessed visually.

8.4 Results

8.4.1 Invasion of the genioglossi and lateral extrinsic muscle group

Based on the visual inspection, 53 (60.9%) of 87 cases were found to exhibit invasion of the lateral extrinsic muscles. 9 (11.7%) of 87 cases were found to have frank invasion of the genioglossi. Of the remaining 78 cases, 60 (76.9%) displayed no invasion while 19 (23.1%) were thought to reach but not invade the genioglossi. Extrinsic muscle invasion did not predict OCLNM. However, lateral muscle invasion predicted ACS and DFS. Detailed results are shown in Table 8.2 and Figure 8.8.

8.4.2 TNM reclassification based on fusion results

When the strict definition of a cT4 tumour was applied to the cases based on atlas-identified lateral extrinsic muscles invasion 37 cases were upstaged to cT4. This upstaging included 8 of 31 cT1, 26 of

Figure 8.7: MRI-atlas fusion. A screen shot of the FantaMorph GUI during an MRI-atlas fusion. In the top left, the selected atlas is shown with the tongue contour outlined in red and the extrinsic muscles outlined in green. The top right image shows a coronal T2-weighted MRI FSE of a patients with a T2 SCC of the left lateral tongue. The green dots represent control points assigned to the tongue contour, facilitating the deformation of the atlas image. The final product is shown in the lower image, the deformed atlas overlaid onto the MRI image. Note that the tumour appears to be invading the lateral intrinsic muscle pair.

35 cT2 and 3 of 5 cT3 cases. Table 8.1 compares the pre and post-atlas cT staging.

Table 8.1: Clinical T classification before and after atlas application. A large number (37) of cT1, cT2 and cT3 cases are upstaged to cT4.

	Pre-atlas	Post-atlas
cT1	31	23
cT2	35	9
cT3	5	2
cT4	16	53

Table 8.2: Results of extrinsic muscle invasion for prediction and prognosis. Neither lateral extrinsic muscle or genioglossus invasion was predictive of OCLNM. Lateral muscle invasion predicted ACS and DFS, but not DRS. Genioglossus invasion was of no prognostic value. OR: odds ratio; CI: confidence interval; OCLNM: occult cervical lymph node metastasis; ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival; HR: hazard ratio. Significant results are indicated in green.

	OR	*95% CI*	*p-value*
OCLNM			
Lateral muscle invasion	*2.08*	*(0.58 - 7.74)*	*0.276*
Genioglossus invasion	*2.15*	*(0.54 - 8.58)*	*0.278*
	HR	*95% CI*	*p-value*
ACS			
Lateral muscle invasion	*4.47*	*(1.01 - 19.79)*	*0.049*
Genioglossus invasion	*1.75*	*(0.63 - 4.82)*	*0.283*
DRS			
Lateral muscle invasion	*3.46*	*(0.76 - 15.79)*	*0.109*
Genioglossus invasion	*1.96*	*(0.631 - 6.08)*	*0.245*
DFS			
Lateral muscle invasion	*3.66*	*(1.06 - 12.65)*	*0.040*
Genioglossus invasion	*1.29*	*(0.50 - 3.33)*	*0.597*

8.4.3 Prediction of cervical lymph node metastasis based on fusion results

The ability of pre and post-atlas clinical T to predict pathological CLNM is shown in Table 8.3. In the post-atlas scenario, the majority of cases positive for CLNM are found in the cT4 grouping. However, the clinical utility of the cT classification for assessing the cervical lymph nodes is negated by the fact that nearly all cases were classified as cT4.

Table 8.3: Pathological lymph node status by pre-atlas and post-atlas clinical T. When the lateral extrinsic muscle invasion is used to classify patients as cT4, the majority of positive cases occur in the cT4 group. However, the ability of cT to discriminate between positive and negative cases is reduced as nearly all of the cases have been designated cT4 (77.8%).

	Pre-atlas		*Post-atlas*	
	Negative	*Positive*	*Negative*	*Positive*
cT1	*25*	*5*	*19*	*3*
cT2	*23*	*12*	*7*	*2*
cT3	*3*	*2*	*1*	*1*
cT4	*7*	*9*	*31*	*22*

8.4.4 Prediction of survival based on fusion results

The ability of pre and post-atlas cT to predict survival is shown in Table 8.4. In the post-atlas scenario, the majority of survival events are found in the cT4 category. However, the clinical utility of cT for prognosis is negated by the fact that now nearly all cases are classified as cT4. Survival is diagrammed in Figure 8.8.

Table 8.4: Survival by pre-atlas and post-atlas clinical T. When the lateral extrinsic muscle invasion is used to classify patients as cT4, the majority of events occur in the cT4 group. However, the ability of cT to discriminate between different prognostic groups is reduced as nearly all of the cases have been designated cT4 (77.8%).

	ACS Events		DRS Events		DFS Events	
	Pre-atlas	Post-atlas	Pre-atlas	Post-atlas	Pre-atlas	Post-atlas
cT1	2/30	1/23	2/31	1/23	6/30	2/23
cT2	7/35	0/9	4/35	0/9	5/35	0/9
cT3	1/5	1/2	1/5	1/2	1/5	1/2
cT4	5/16	13/53	5/16	10/53	6/16	15/53

8.5 Discussion

8.5.1 The tongue atlas leads to TNM upstaging

If the current TNM definition of "extrinsic muscle" was used as stated in the most current edition of the TNM staging criteria, a large number of tumours would be upstaged. In our sample, 37 of 71 (52.1%) cases clinically staged as T1, T2 or T3 were upstaged to cT4. This reflects the fact that the vast majority of lingual SCCs are found in the lateral tongue, where a depth of only millimeters will

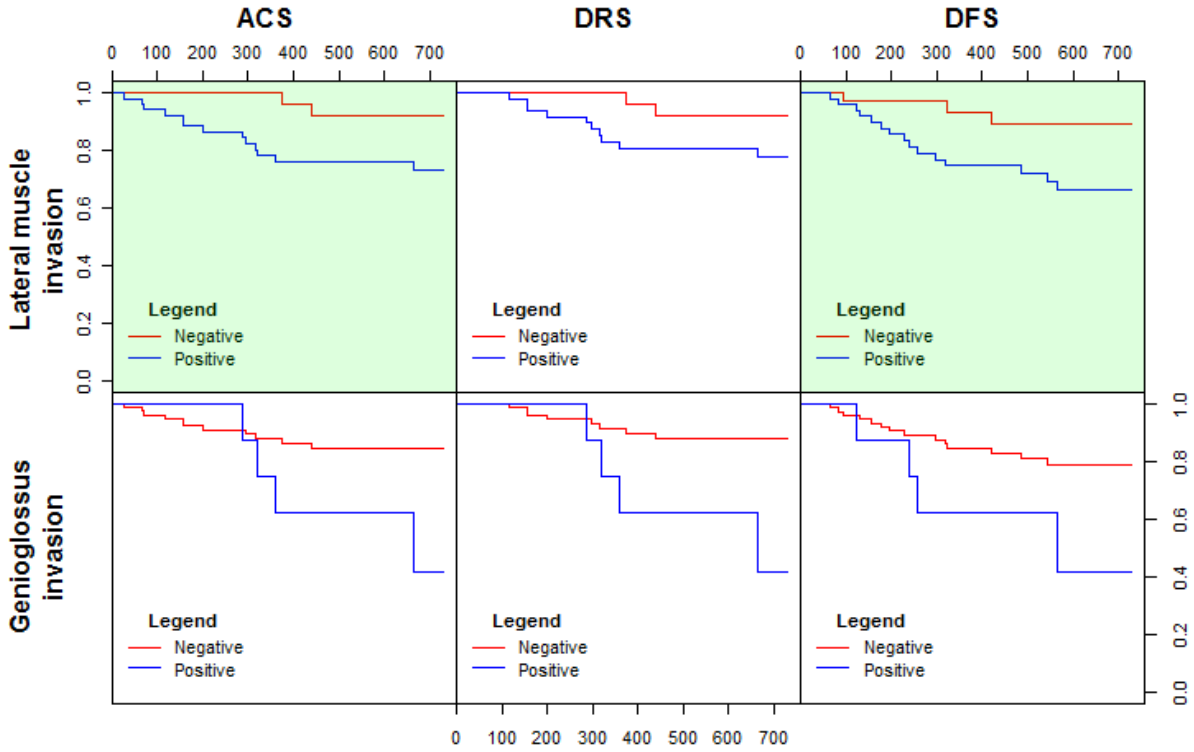


Figure 8.8: Results of extrinsic muscle invasion for prognosis. Lateral extrinsic muscle invasion predicted ACS and DFS, but not DRS. Genioglossus invasion did not predict survival. ACS: all-cause survival; DRS: disease-related survival; DFS: disease-free survival; HR: hazard ratio. Significant results are indicated by green shading.

lead to lateral muscle invasion. This also reflects the fact that radiologists and surgeons do not consider the lateral extrinsic muscles when staging lingual SCC despite obvious instructions to do so as given by the AJCC.

8.5.2 Predictive and prognostic value of extrinsic muscle invasion

Statistical analysis showed that lateral extrinsic muscle invasion did not predict OCLNM. However when survival was analyzed, lateral muscle invasion did predict ACS and DFS, but not DRS. In theory, those cases negative for lateral extrinsic muscle invasion should have improved survival and a resulted in a lower risk for OCLNM, only because this group would have superficial tumours with very little invasive depth. This was seen for ACS and DFS, but not OCLNM or DRS.

One explanation is that many tumours positive for lateral muscle invasion are still very superficial. As such, the "positive for invasion" group becomes a mixture of tumours ranging from superficial to very deep. The mixed nature of the "positive grouping" lowers the rate of OCLNM and raises the median survival of the included cases. Thus, the relationship between lateral muscle invasion and OCLNM or survival is less distinct and not seen when the small sample sizes of this study are used.

There was no relationship between genioglossus invasion and OCLNM or survival. This is likely due to the small number of cases that actually exhibit genioglossus invasion (9 cases).

8.5.3 Practical interpretation of T4 criteria

There are two reasons why a surgeon or radiologist may choose to ignore lateral muscle invasion when staging lingual carcinoma:

> *1) The healthcare professional may not be aware of the superficial position of the styloglossus and hyoglossus and therefore do not classify the tumour as invading extrinsic muscle.*
>
> *2) Based on experience and anecdotal evidence, lateral extrinsic muscle invasion is ignored as it is a poor prognostic factor and of little use for directing surgical management.*

The former reason is likely to be only a small part of the explanation, given the extensive knowledge

of anatomy possessed by surgeons and radiologists specializing in OCSCC. Therefore, the second explanation remains.

A cT4 tumour receives at least a stage grouping of IV. Such a stage grouping indicates that the patient is at increased risk of disease-related mortality and guides the surgeon to adopt aggressive therapy. Practically speaking, a tumour that is classified as cT4, based solely on invasion of the lateral extrinsic muscles, is unlikely to be extensive. Second, such tumours are highly accessible to the surgeon and clear margins can be achieved with minimal difficulty.

8.5.4 Recommendations for modification of T4 criteria

Using the VHF atlas, invasion of the lateral extrinsic muscles did have some value for prediction of survival. Thus, its inclusion as a T4a staging criteria may be justified. However, if the definition of a T4a tumour was interpreted to include lateral muscle invasion, there would be a staggering amount of superficial tumours classified as stage IV. Stage IV tumours are associated with decreased survival and warrant aggressive treatment that includes adjuvant therapy. Stage IV tumours, designated as such solely on the basis of superficial lateral extrinsic muscle invasion, would be massively over-treated, causing undue iatrogenic morbidity and mortality. Thus, while lateral extrinsic muscle invasion is statistically relevant, clinical practicality is a limiting factor.

This supports the hypothesis that tumours invading the lateral extrinsic muscles are inappropriate for classification as cT4 tumours. A more practical approach is to limit the criterion for cT4 classification to the genioglossus and/or the midline. These extensive tumours are associated with particularly high mortality and require aggressive management.

A more complete discussion of the implications to TNM staging of the oral cavity, as well as suggestions for TNM modification, is found in Chapter 9.

8.5.5 Comments on the VHF tongue atlas

The VHF atlas was developed through previous work in the Department. It was designed using a

computer science and engineering approach, though clinical utility was a consideration. As a result, there are several aspects of its design that must be considered.

8.5.5.1 The VHF tongue protrudes at the apex and laterally

This shortcoming was discussed in Section 8.3.2 of the methods, but deserves to be mentioned again. An ideal tongue atlas would reflect a more anatomical tongue position.

8.5.5.2 The components of the VHF atlas

There are some problems with the identification of the anatomic components of the VHF atlas. The outline of the superior tongue is relatively easy to define. However, a choice must be made on how to define the inferior limit of the tongue. For the VHF atlas, the superior surface of the mylohyoid was used as the inferior limit. However, the sublingual glands were not included. Thus, the inferior limit is best described as the superior surface of the sublingual glands laterally and the superior surface of the mylohyoid medially. Beyond the posterior edge of mylohyoid, the inferior outline of the atlas appears to be limited by a line connecting the coronal cross sections of the greater cornua of the hyoid bone.

The VHF atlas includes the geniohyoid in addition to the extrinsic and intrinsic muscles of the tongue. The anterior limit of the VHF atlas is determined by the tip of the tongue. The posterior limit seems to have been chosen arbitrarily and does not include any of the lymphoid tissue of the tongue base.

8.5.5.3 The complex shape of the genioglossus

The genioglossus has been separated into two units. The main genioglossus is pyramidal in shape in the posterior and anterior thirds of the tongue. However, in the middle third the shape is complex and appears as distinct upper and lower muscle bundles separated by a central "stalk". The secondary genioglossus is circular in shape in the posterior tongue and assumes a more ovoid shape as one moves anteriorly through the atlas. Review of the VHF coronal views suggests that several well defined fibrofatty planes within the substance of the genioglossi are responsible for their complex appearance.

8.5.5.4 The hyoglossus and styloglossus are merged

No attempt is made to distinguish between the hyoglossus and styloglossus. Instead, a single lateral extrinsic muscle is given in the atlas. This lateral muscle represents either the combined muscle bulk of the styloglossus and hyoglossus or only the hyoglossus. The former option is the more likely of the two as the lateral muscle bundle is seen to abut the outline of the tongue as defined by the atlas. Combination of the lateral extrinsic muscles does not affect the utility of the atlas for MRI co-registration as separation of the two is beyond currently available spatial resolution.

8.5.5.5 Delimiting the extrinsic muscles: an arbitrary decision

While the extrinsic muscles of the tongue have well defined bony origins, the intermingling of their fibres with those of the intrinsic muscles is much more gradual. Thus, it is difficult to define a point where the extrinsic muscles end and the intrinsic muscles begin. The VHF atlas took a conservative approach to defining the insertion of the extrinsic muscles that, in my opinion, could have been extended.

This is noticeable in the lateral extrinsic muscles. Classical descriptions of the styloglossus indicate that it extends to the apex of the tongue [223-224]. DWI tractography suggests that the styloglossus intermingles with the inferior longitudinalis in the anterior one-third of the tongue before reaching the apex [229]. The present observations of the VHF support the DWI tractography conclusions. In the VHF atlas, the lateral extrinsic muscle group terminates in the middle one-third of the tongue. For the reasons mentioned in the previous paragraph, extension into the anterior one-third of the tongue and perhaps even to the apex, would give a more accurate description.

8.5.5.6 The palatoglossus is not included

The palatoglossus is not shown as part of the atlas, due to its small size.

8.5.5.7 The atlas excludes extra-lingual landmarks

The VHF atlas does not include any oral cavity structures outside the tongue. Extralingual structures, such as the mandible, maxilla, sublingual glands and sinuses, are useful landmarks for interpretation.

Such landmarks may also be useful for future attempts at automated registration techniques but would also greatly increase the information content of the atlas.

8.5.6 The clinical implementation of VHF atlas

The implementation of the VHF atlas requires considerable input from a trained operator. Fusion times vary, depending on the extent of the tumour in question. If it is assumed that 20 control points are used to fuse the atlas with 5 tumour containing slices, a time of 8-10 minutes can be expected. While the learning curve for atlas fusions is minimal, knowledge of oral cavity anatomy, as well as familiarity with both the atlas and oral cavity MRI, is prerequisite.

While the time required to implement the atlas is prohibitive, the small number of lingual carcinoma cases and the valuable information offered by the result of the fusion process favour its use. Use of the atlas has the potential to make surgeons more aware of the position of the extrinsic muscles, in particular the styloglossus and hyoglossus. This has implications for TNM staging, especially if strict interpretation of the T4a classification criteria is assumed. In either case, knowledge of lateral extrinsic muscle involvement and the extent of the surgical defect is essential for predicting post-surgical deficits in tongue function.

8.5.7 Shortcomings of MRI-atlas fusion

8.5.7.1 The selection of the atlas slice

The selection of the atlas slice was made difficult by the protrusion and winging of the tongue in the VHF. As a result, several pre-processing steps had to be carried out to make the atlas usable. Even then, a standard scaling approach sometimes produced results in which the suggested atlas slice was not matched to the MRI slice. While an attempt was made to maintain a standardized approach and use the suggested slice, sometimes it was necessary to choose an atlas slice subjectively.

One option that would overcome this challenge is to use a three-dimensional co-registration instead of the current two-dimensional technique. However, there is no pre-existing software for this computationally difficult task, which is beyond the scope of the present work.

8.5.7.2 Difficulty in the posterior tongue

Because the atlas does not include the lymphoid tissue in the posterior tongue, the problems associated with selecting the appropriate slice are compounded. The length of the tongue is measured from the posterior limit, as seen on T2-weighted MRI, which includes the lymphoid tissue.

8.5.7.3 Fusion using poor quality MRI

Both motion and susceptibility artefact obscure the outline of the tongue. However, considering that any tongue-atlas fusion is a approximate estimate of extrinsic muscle position, fusion can still be attempted in the presence of mild or even moderate artefact with little appreciable loss of accuracy. Nevertheless, 9 studies were excluded because of poor quality coronal studies.

8.5.7.4 The atlas is based on a single tongue

The atlas is based on a single tongue. It may be that the atlas is not representative of the most prevalent lingual anatomy present in the population. However, agreement with classical descriptions is reassuring. Even if the atlas represents the most common arrangement of lingual myoarchitecture, population variation means that the atlas will still be inaccurate in a significant number of cases.

8.5.7.5 The atlas is a rough guide to extrinsic muscle position

VHF atlas fusion with MRI is one of few studies to show the relative position of the extrinsic muscles. However, the results must be viewed with caution. The VHF atlas provides an approximate indication of the location of the extrinsic tongue muscles. There are several reasons for this:

> *1) The VHF atlas is based on the anatomy of a single person. It is likely that oral and lingual anatomy varies widely from person to person.*
>
> *2) The distortion of the lingual contour produced by the post-mortem position of the VHF makes co-registration with MRI series difficult. This is especially true in the lateral tongue, which is "winged" as a result of the largely edentulous VHF.*
>
> *3) Co-registration is a manual process requiring the operator to select a series of control points. The process is therefore operator dependant and, to some extent, arbitrary.*
>
> *4) Atlas slice selection is complicated by adjustments made to compensate for tongue protrusion in the VHF. As a result, the standardized approach to slice selection adopted by*

this work can, at times, lead to the selection of atlas slices that represent lingual anatomy more posterior or more anterior to the MRI slice position.

While it is still essential to adopt a standardized approach to atlas co-registration, the position of the extrinsic muscles, particularly the lateral extrinsic muscles, must be interpreted as an approximation.

8.5.7.6 The need for semi-automated or automated co-registration

While manual co-registration allows implementation of the atlas, the process is inefficient. Semi-automation of the co-registration process would remove many operator-dependent elements while overcoming some of the complexities of full automation. Work on semi-automation has been carried out as part of the previous work on the atlas. However, at present, the semi-automation routine cannot be implemented retrospectively due to the inability of the algorithm to distinguish the lingual contour from adjacent structures. There is an argument for a modification of the MRI protocol that requires patients to coat the mucosa of the oral cavity with a Gadolinium-containing pectin gel.

8.6 Conclusions

The VHF atlas, when fused with coronal MRI images, is a useful guide to highlight the invasion of the extrinsic muscles of the tongue. It must be kept in mind that the atlas fusion is a rough approximation and should be thought as a qualitative evaluation. A more rigorous application of atlas fusion lies in the realm of computer science and is beyond the scope of this work.

Lateral extrinsic muscle invasion is not considered during clinical staging, though they are included in the TNM definition of a T4a tumour. Whether this is by omission or design is unclear. However, when the atlas is used to evaluate extrinsic muscle invasion, inclusion of lateral extrinsic muscle invasion leads to upstaging a nearly 55% of cT1, cT2 and cT3 SCC of oral tongue tumours. This upstaging is inappropriate and does not reflect decreased survival or a need for more extensive therapy.

The appropriate step to address this is a modification of TNM criteria to restrict extrinsic muscle invasion to only the genioglossus or crossing of the midline. Though genioglossus invasion was not shown to predict OCLNM or survival in this work, the small sample size likely limited statistical

power. However, in Chapter 5 midline invasion was shown to be of prognostic and predictive value. Whatever the case, inclusion of styloglossus and genioglossus invasion as criteria for T4a designation is inappropriate and needs to be excluded or qualified by the AJCC.

9 General discussion and conclusions

9.1 Introduction

This work is a detailed study of the role of tumour morphometry in the natural history, disease staging and prognosis of squamous cell carcinoma of the oral cavity. The approach was multifaceted but centered around the development of a detailed, multicenter database of OCSCC patients. MRI was used to measure tumour volume, thickness and maximum diameter on 199 cases. There are no other studies of MRI-based measurement to achieve this sample size and statistical power.

First, a critical review of the depth of invasion literature was carried out with an emphasis on the role of statistical methodology in cut-point variability. The results were supported by statistical simulation using real OCSCC patient data. Second, a detailed analysis of the information gathered in the database was carried out. A number of patient and tumour factors were identified as predictors of occult cervical lymph node metastasis and survival. These factors were included in multifactor models built using regression, artificial neural networks and recursive partitioning. Finally, the extrinsic muscles of the tongue, a controversial component of TNM staging, were studied using the Visible Human Female dataset and the distance of the styloglossus and hyoglossus from the lingual surface measured. An anatomic atlas of the tongue was fused with clinical MRI of OCSCC patients and the role of lateral extrinsic muscle invasion in TNM staging challenged. This thesis thus makes contribution to the field of oral cancer staging and prognosis.

In the following section, this contribution is outlined in detail. Shortcomings of the thesis are then discussed. A detailed section on continuation of the present work follows. The final section summarizes the conclusions of the work.

9.2 Contribution of this thesis to oral cancer research

9.2.1 Design and implementation of comprehensive OCSCC database

Oral cancer is a rare disease requiring extensive, resource-intensive treatment. The number of cases at any one institution is limited. This means that building a database based on retrospective chart

review requires the architect to identify potential candidates from among a heterogeneous group of cases. In a setting where paper-based records are in widespread use, case identification and retrieval can be an arduous task. In this work, the final sample of 199 cases was the result of distillation from thousands of potential candidates.

The sample size is one of the strengths of this work. In the review of the depth of invasion literature, the median sample sizes of work studying CLNM or survival were 50 and 90 cases respectively. If one considers the extremely stringent inclusion and exclusion criteria, the significance of the sample size is apparent. The published results of Chong et al. on lingual SCC volume of invasion, one of the few to explore MRI volume, consisted of only 17 cases [240].

Retrospective case review is highly dependent on the quality of information contained in the patient records available at participating institutions. Thus, database design is dictated by the available information. The result is that the database used in this study is very specific to this work. Despite this, the database contains a wealth of information and is a solid foundation for further work at the participating institutions.

9.2.2 Statistical methods are a source of depth of invasion cut-point variability

Depth of invasion has emerged as a candidate for refinement of the TNM staging criteria for OCSCC. There seems to be a consensus that depth is a highly significant predictor of OCLNM and survival. Disagreement exists, however, as to the proposed cut-point (or cut-points) to be used to stratify risk.

This work is the first to investigate and identify statistical methodology as a potential source of cut-point variability. Specifically, the categorization of continuous data is a poor statistical practice that results in the loss of valuable information. When categorization is coupled with the use of the "lowest p-value" method, the results can be misleading. Statistical simulation showed that the cut-point value chosen using this method is data dependent and small variations in sample composition can result in very different suggested cut-points. OCSCC sample sizes are small and, thus, the potential for variability high.

Perhaps more disturbing is the fact that p-values reported in these circumstances can be spuriously significant. This work has shown, through simulation, that even when no relationship exists between the dependent and independent variable, 30% of tests result in p-values less than 0.05. The implication is that a large portion of the OCSCC depth of invasion literature may be based on falsely-significant results. Considering that there is a tremendous push to include the depth of tumour invasion in TNM staging criteria, it may be advisable to reassess the body of evidence.

9.2.3 Many patient and tumour factors contributed to OCLNM and survival prediction

Chapter 5 of this work evaluated the predictive and prognostic ability of each individual patient and tumour factor. The result is a library of information detailing each element of the information contained in the database.

Many of the factors evaluated in Chapter 5 have been studied in the literature previously. However, like other OCSCC work, many of the sample sizes used to study these factors were small. The sample size of the present study, 199, is large and offers a valuable contribution to the study of oral cavity cancer.

Similarly, comparison of the results of the univariate analysis with previous work allows the sample used in the present work to be characterized. Any unique elements of the data can be highlighted and investigated. One example is perineural invasion. PNI has been noted by several authors to be predictive of OCLNM and survival. In this work, PNI only predicted all-cause survival. This might indicate that the unique composition of the sample constructed for this work behaves differently than that of other studies. Thus, results must be interpreted in context.

Finally, evaluation of each individual factor identified candidates for multivariate modeling. This was a necessary step to constrain the complex task of modeling a large number of factors with a relatively small number of cases.

9.2.4 Volume is at least equivalent to thickness for OCLNM prediction

Many factors were identified as potential predictors of OCLNM by univariate analysis. Ideally, a predictive model would include many of these factors in order to offer the best discrimination between OCLNM probabilities for individual cases. However, only 136 of the 199 cases were clinically N0. Further, only 25 of these cases were positive for CLNM on pathology. Unfortunately, model complexity is limited by the number of cases in the smallest category.

Using the pre-surgical variable pool, the volume of invasion played an important role in the top-ranked logistic regression, ANN and recursive partitioning models. Though imaging bone invasion was included in these models, its contribution was questionable. In the Cox model, the odds ratio of bone invasion was 0.33 (p=0.532). Though insignificant, indicates that the presence of SCC invading bone may actually lower the risk of metastasis. This is compared to an OR of 2.91 on univariate analysis. Similarly, recursive partitioning included only volume in the final decision tree.

While MRI thickness (11.0 mm) was equivalent to volume in all three top-ranked pre-surgical models, the results of post-surgical modeling suggest that volume is superior. Further work, with a larger sample size, is required to definitively differentiate between the two.

MRI-based tumour volume measurement is a novel predictive factor that has not been fully explored in the OCSCC literature. Specifically, the top-ranked OCLNM models used MRI volume categorized using a 3.0 cm^3 cut-point. These models were able to predict OCLNM with a 95% negative predictive value. Published figures state that surgeons may adopt a "watch and wait" approach when the probability of OCLNM is below 20% [3]. Volume of invasion has the potential to identify patients with probabilities of OCLNM below 5%. Surgeons are likely to find this risk figure far more acceptable.

9.2.5 Artificial neural networks offer superior, individualized prognostic information

Traditionally, Cox proportional hazards models have been used for survival analysis of censored data. While ANNs have been compared to Cox models for use in other oncology disciplines, the results have been mixed. When pre-surgical and post-surgical survival were modeled using the variables

identified in univariate analysis, ANNs emerged as the superior method.

Pre-surgically, the top-ranked ANNs for ACS and DRS were composed of MRI-based volume, clinical N-stage and midline invasion. The results were very similar for post-surgical modeling, with volume remaining and pN replacing cN. There is, thus, a case for a unified ANN model of survival using volume, N-stage and midline invasion.

What is remarkable about these results is that they reflect the components of the T and N criteria of the TNM system. Volume is a useful substitute for maximum diameter. While the continuous form of volume was used in the top-ranked models, the results of univariate analysis suggest that categorization using the mean or quartiles would provide good discrimination between prognostic groups. Midline invasion is an anatomy-based factor, not unlike bone or extrinsic muscle invasion, currently used as part of the T4a criteria. It is not surprising that cN and pN are vital components of the final models, given their role as strong univariate predictors of survival.

Depth has been widely acknowledged as a replacement for diameter in TNM criteria, yet was not superior to volume in this work. It has already been established that poor statistical practices make much of the depth of invasion literature open to question. Further, many studies test depth of invasion only as a univariate predictor of survival. If depth is indeed a strong predictor of CLNM, inclusion in a multivariate model with cN or pN could negate some of its univariate value. If volume offers superior information on local spread and the risk of local recurrence, it could result in a better model for survival prediction.

There are few, if any, other works in the OCSCC literature that use ANNs for survival prediction. This is likely a result of the small sample sizes that characterize most OCSCC studies. The traditional TNM approach stratifies patients into groups based on broad criteria. Using the ANN, surgeons can tailor prognosis to each individual patient based on their unique set of explanatory variables. Future work with a larger sample size would include more explanatory variables and the prognosis derived even more patient focused.

9.2.6 Comparison of volume, diameter and thickness

Based on the univariate statistical analysis it is difficult to draw any conclusions about the relative value of MRI-based volume, diameter and thickness. All three factors had similar odds ratios and hazard ratios when treated as continuous, log transformed or categorized variables.

When used to build multivariate models, the differences between measurement-based predictive factors were much more pronounced. MRI-based volume of invasion, categorized using the median (3.0 $cm^{3)}$, was clearly superior to MRI-based diameter and thickness for the post-surgical prediction of OCLNM. In all three modeling methods, volume (3.0 cm^3) was the central component of the top-ranked result, with a NPV exceeding 95%. No top-ranked model used depth or diameter.

The relationship is less clear cut on pre-surgical OCLNM analysis. Volume and thickness had equivalent NPV scores with all three modeling methods. All of the pre-surgical models were simplistic and likely dominated by the measurement variable. A larger sample size would allow the modeling methods to better evaluate complex relationship between variables, and better distinguish volume and thickness. However, as mentioned, the results of post-surgical modeling suggest that volume may be superior to thickness.

ANNs were the most useful predictors of both pre-surgical and post-surgical ACS and DRS. Pre-surgically, all ANNs included MRI-based volume. Post-surgically, volume remained in all ANN models. The only exceptions were the ANNs for DFS prediction. These models both used diameter, suggesting that diameter may be superior to volume for recurrence prediction. It was also unclear if volume as a continuous variable or log transformed volume was superior. Further work with larger sample sizes is required to clarify this dilemma.

The selection of the most useful measurement-based factor should take into account clinical practicality. The principal benefit of MRI-based thickness and diameter is that they can be measured quickly using existing MRI protocols. Linear measurement is also a familiar task to radiologists, and all modern commercial DICOM viewing software contains the appropriate applications. Similarly, depth

and diameter are routinely and easily measured on post-surgical specimens. The adoption of predictive models based on linear measurement does not require the introduction of new techniques.

Volume measurement is a more cumbersome task. Specialized software, not routinely included in commercial packages, is required. As well, volume measurement is a relatively time-consuming task and inclusion in routine clinical practice requires significant justification. Compounding this is the fact that pathological volume measurement requires very specialized techniques and equipment.

The principal benefit of volume has to do with the complex shape of the oral cavity. Tumours of the oral sulci and alveolar ridges often have complex surface contours. Consider a tumour of the lower gingivobuccal sulcus. Anatomically, the mucosal surface of the sulcus is reflected back on itself. On MRI, the juxtaposed mucosal surfaces are indistinguishable from the tumour bulk. As a result, depth and diameter must be measured from an estimated mucosal surface. In this work, the estimated mucosal surface was taken to be the midline contour of the tumour. Volume does not require the identification of the mucosal surface and thus provides objective data in these circumstances. Another example occurs in cases of medullary bone invasion. If the tumour follows the marrow cavity, the shape can be complex indeed. The "point of deepest invasion" in these cases is difficult to identify but need not be identified for volume. This is diagrammed in Figure 9.1.

The ability of volume to overcome these obstacles would likely result in improved interobserver variability compared to depth and diameter. This is of key importance for any staging protocol that is to undergo widespread use. Intraobserver variability was not evaluated empirically in this work. However, the work of Chong et al. has shown that volumes measured for tumours of the oral tongue by different observers do not differ significantly [240]. The group also demonstrated that intraobserver variability could be improved by introducing a semi-automated measurement procedure [240].

Figure 9.1: A coronal diagram of a tumour of the right gingivobuccal sulcus. The tumour is highlighted in red, while the diameter and depth are shown in yellow and blue respectively. In this diagram, the oral cavity is filled with air and the contour of the mucosal surface easily seen. When MRI takes place, the oral cavity becomes a potential space and the contour of the mucosal surface is no longer apparent. A tumour like the one pictured would appear as a singular, globular mass and the diameter and depth would be difficult to measure accurately. Adapted from: [241].

Given the results of this work, it is my opinion that the benefits of volume of invasion measurement exceed its shortcomings. For ease of clinical adoption, the MRI-based volume is the better option for survival prediction using multivariate predictive models. The one exception is the prediction of recurrence, in which diameter may still have a role.

As MRI resolution increases and semi-automated, volume-measuring software is developed the case for MRI-based volume of measurement will strengthen. As Radiology moves towards computer-based image processing, cancer prognosis will move towards volume.

9.2.7 "[Deep] extrinsic muscle" invasion must be reconsidered as a T4 criterion

Chapter 7 of this work has shown that the lateral extrinsic muscles of the tongue are located very close to the surface mucosa. Since the majority of oral tongue SCCs are located in the lateral tongue, a large number of tumours invade the extrinsic muscles. This was confirmed by the work in Chapter 8 using the VHF-based anatomic atlas.

The implication of this finding is that very small, superficial tumours of the lateral tongue are likely to be invading extrinsic muscle. Technically, these tumours should be classified as T4a and, therefore, Stage IV. Yet, these tumours are unlikely to justify the poor prognosis and aggressive treatment that accompanies Stage IV disease.

Despite this, it was evident from reviewing patient reports that it is rare for a superficial tumour of the lateral tongue to be classified as T4a. Only when tumours were extensive and either invading the genioglossus or crossing the midline were such a designation considered. It is likely that surgical experience and practicality prevent strict application of T4a criteria. The current T4a criterion of "deep [extrinsic] muscle of the tongue [genioglossus, hyoglossus, palatoglossus, and styloglossus]" is, therefore, not an appropriate reflection of current practice and should be reconsidered.

9.3 Proposals for refinement of TNM staging criteria

One of the main objectives of this thesis was to evaluate and propose changes to controversial

aspects of the TNM staging system for oral cavity cancer. Specifically, the criteria in question are diameter-based classification of T1, T2 and T3 tumours, as well as extrinsic muscle invasion-based classification for T4a tumours.

9.3.1 Replace diameter with volume of invasion

The diameter has been in widespread use by the oral cavity TNM staging system for more than 40 years. It is entrenched in surgical practice and cannot be changed without good reason. Doing so would constitute a revision of the primary structure of TNM staging of oral cavity cancer and make comparison with all previous work difficult.

There is evidence to suggest that MRI-based volume of invasion is a candidate to replace maximum diameter. As discussed in Section 9.2.2, while depth is a significant predictor of OCLNM and survival, there is widespread disagreement as to the appropriate cut-points. The depth literature is characterized by small sample sizes and sample heterogeneity, and lacks the focus to provide convincing evidence for TNM reform. This work has identified poor statistical methodology as a source of cut-point variability and false-positive results, calling into question a large portion of the depth literature.

This work supports the role of volume as a criterion for TNM staging. Multivariate modeling has shown that volume was a component of top-ranked pre-surgical and post-surgical models. While volume was used as a continuous variable in the top-ranked models, univariate analysis demonstrated that categorization using the median or quartiles provides good definition between prognostic groups. Following the two cut-point convention of the TNM staging system, the quartile cut-points, 0.5 cm^3 and 8.0 cm^3 would be appropriate.

9.3.2 Introduce a T4a category specific to oral tongue

Change is also required for the extrinsic muscle-based T4a criterion. At the moment, invasion of the lateral extrinsic muscles is not considered by surgeons and, thus, change would not significantly impact current practice or prevent comparison with previous work. It would, however, prevent

confusion and remove an objective part of T4a classification.

The styloglossus and hyoglossus muscles are located superficially and therefore invasion does not warrant T4a designation. The palatoglossus is a very small muscle, difficult to identify and, anatomically, not part of the oral cavity. The genioglossus is located deep in the oral tongue and invading tumours are extensive. However, the deep location of the genioglossus is not universal. In the floor of the mouth, the genioglossus is found just beneath the surface mucosa. Invasion by small, superficial tumours is possible. Therefore, invasion of the extrinsic muscles should not be used as a T4a criterion.

Like the lip, the oral tongue should be given its own T4a criteria. The number of oral tongue tumours is comparable to the number of lip tumours. Specifically, the new T4a (oral tongue) grouping should identify invasion of the genioglossus as a criterion. Provisionally, extension to the floor of the mouth should also be considered as a T4a (oral tongue) criterion. Though this is not supported by this work, it is an extension of the principle that extrinsic muscle invasion is a route of extension to the floor of the mouth.

Accompanying the creation of the T4a (oral tongue) category would be the removal of "[Deep] extrinsic muscle" from the T4a (oral cavity) criteria. Doing so ensures that floor of the mouth tumours invading the genioglossus or hyoglossus are not designated as T4a.

9.3.3 Summary of proposed changes

In summary, the revised T4 criteria for tumours of the oral cavity would become:

Table 9.1: Proposed changes to TNM criteria

T1	*Tumour 0.5 cm^3 or less in greatest dimension*
T2	*Tumour more than 0.5 cm^3 but not more than 8.0 cm^3 in greatest dimension*
T3	*Tumour more than 8.0 cm^3 in greatest dimension*
T4a (lip)	*Tumour invades through cortical bone, inferior alveolar nerve, floor of mouth, or skin of face, i.e., chin or nose*

T4a (oral tongue)	*Tumour invades adjacent structures (e.g., through cortical bone, the genioglossus, floor of mouth)*
T4a (oral cavity)	*Tumour invades adjacent structures (e.g., through cortical bone, maxillary sinus, skin of face)*
T4b	*Tumour invades masticator space, pterygoid plates, or skull base and/or encases internal carotid artery*

9.3.4 Consider an individualized approach to OCLNM prediction and prognosis

TNM staging criteria are firmly entrenched in OCSCC prognosis. The case for change must overcome considerable inertia. When change does take place, it improves two of the five central tenets of TNM staging:

The selection of primary and adjuvant therapy
Estimation of prognosis

However, this improvement comes at the permanent, or at least temporary, cost of the three remaining tenets:

Assistance in evaluation of the results of treatment
Facilitation of the exchange of information among information centers
Contribution to the continuing investigation of human cancers

This is the central dilemma of TNM modification and not to be taken lightly, especially in the case of oral cancer where building a sufficient sample size can take decades.

Every day evidence-based medicine challenges oncology treatment and prognosis paradigms with new surgical techniques, novel pharmaceuticals and an increasingly in-depth understanding of the molecular basis of disease. TNM staging criteria are slow to change and threaten to be left behind.

Oral cancer treatment and prognosis must move beyond the TNM system for the selection of therapy and the estimation of prognosis. As this work demonstrates, this can be accomplished on an individualized basis using multivariate models. These models can be adapted to new research or emergent prognostic factors in such a way that they keep pace with the dynamic surgical

environment.

Nevertheless, the TNM criteria should remain. However, their focus should only be the exchange of information among centers and the contribution to continuing research. The current TNM criteria are basic enough to be relevant in less technologically advanced surgical centers. Thus, researchers can continue to have a productive dialogue that transcends borders. Keeping the current TNM criteria also allows for comparison with past work, and maintains a pool of cases for retrospective review.

9.4 Limitations of the work

A number of shortcomings of this work have been discussed in the previous chapters. This section of the discussion will address those limitations that are common to the entire thesis.

9.4.1 The difficulties of retrospective review

The difficulties of retrospective chart review have already been discussed in detail in Chapter 3. However, it must be stressed that this method of data gathering has some serious drawbacks. The lack of a straightforward method to identify oral cancer patients, the continued use of paper records and underdeveloped electronic medical records all present obstacles to retrospective review.

9.4.2 The study sample

9.4.2.1 Sample size and heterogeneity

Even though the sample size was 199 cases, which is large among published studies, it remains a small sample for statistical modeling. As a result, the models obtained were of minimal complexity. This is a problem with all oral cancer work and can only be overcome with multicenter studies.

The oral cavity is composed of a number of subsites that differ in natural history and prognosis of disease. However, study of any of these subsites in isolation, with the exception of the oral tongue, would result in sample sizes much too small to be of sufficient statistical power.

9.4.2.2 Bias introduced by exclusion and inclusion criteria

The stringent inclusion and exclusion criteria used in this thesis have advantages and disadvantages.

Those cases meeting the exclusion and inclusion criteria represent a specialized sub-population within the OCSCC patient population. For example, all patients also underwent MRI imaging. This means that patients unfit or unable to undergo MRI were excluded. Further, those individuals with very superficial disease may not have had pre-surgical MRI.

It must be stressed that the results of this work can only be generalized to patients that fall within the selection criteria of this work. Application of the results beyond the sample criteria constitutes extrapolation and would be unreliable in other populations.

9.4.2.3 A flaw in the definition of occult metastasis

Theoretically, patients designated as cN0 and pN0 on pathological analysis of neck dissection specimens could, indeed, harbour neck metastases. Neck dissection is a sampling procedure that does not remove all of the cervical lymph nodes. Similarly, pathological analysis of resected nodes is also a sampling procedure that might not detect microscopic metastases.

If patients underwent post-surgical adjuvant therapy, existing CLNM would be destroyed without being documented. Thus, the incidence of CLNM is likely higher than documented in this work. This issue is impossible to resolve as it would involve denying adjuvant therapy to patients with suitable indications.

9.4.2.4 Variation in cN detection and sample composition at each institution

Care must be taken to ensure that the characteristics of the samples collected from the JRH and UCLH are comparable. One concern is the different protocols present at each institution for pre-surgical investigation of the cervical lymph nodes. UCLH uses ultrasound-guided fine needle aspiration cytology (US-FNAC) to investigate suspicious cervical lymph nodes. Such a technique requires a highly skilled US-FNAC operator and, therefore, is not a standard procedure at all institutions. It is not done at the JRH, thus, the information available for clinical N staging at each institution was slightly different.

Both centers are highly specialized tertiary care facilities. Therefore, the type of case seen at each institution is unlikely to differ significantly. Nevertheless, cases seen at the JRH and UCLH are likely to be more severe, and difficult to treat, than at less specialized institutions. This is important to remember when considering other institutions for future expansion of the sample size.

9.4.3 On the extent of the "lowest p-value" method

The lack of clear methodology about p-value selection allows us to speculate that there is widespread use of the "lowest p-value" method in the depth literature. However, this remains speculation. Despite this, papers that do not discuss their methodology in a clear and transparent manner must be challenged. This is an essential part of peer review.

9.4.4 Problems with MRI-based measurement

9.4.4.1 The measurement of diameter, thickness and volume

Though MRI has superb soft tissue resolution, three-dimensional spatial resolution is limited by large slice thicknesses and slice spacing. The resulting rectangular voxels reduce the accuracy of volume estimation carried out through manual segmentation. MRI-based volume is, thus, an approximation of actual tumour volume that is highly dependent on the Z-axis resolution of the MRI series. The advent of 3T scanners may reduce this shortcoming, allowing radiographers to reduce slice thickness while maintaining a reasonable signal-to-noise ratio.

Most OCSCCs are associated with a cuff of peritumoural edema. The increased water content of the surrounding tissue causes the tumour to appear larger on T2-weighted MRI. The result is a potential overestimation of tumour dimensions.

9.4.4.2 MRI parameter heterogeneity

The scan parameters used for MRI scans in this work were very heterogeneous. Several scanners from participating and referring institutions were involved. Scans also took place over the span of 10 years and reflect the changes in practice that took place in that interval. The result is widespread variability in the type of pulse sequences, field of view, pixel matrix size, slice spacing, slice thickness,

radiofrequency coil, and many additional parameters. Standardization of MRI parameters was not possible due to the retrospective design.

It could be argued that the scan heterogeneity reflects the unpredictability of clinical practice. That the results are compiled under these circumstances is a bonus, and demonstrates the robustness of this work.

9.4.4.3 Measurement was by a single observer

MRI-based measurement in this work was carried out, in duplicate, by the author. Measurements were not carried out by experts in head and neck radiology. However, when required an expert in head and neck radiology (SJG) was available for consultation.

Duplicate measurement by a single author allows for the measurement of intraobserver variability, but not interobserver variability. Interobserver variability is of vital importance if a technique is to be used for multicenter work. It is recommended that interobserver variability be studied prior to further work.

9.4.5 Problems with data analysis

9.4.5.1 Survival time was only two-years

At both institutions, the standard of patient follow-up was five years. However, for some cases with minimal disease the follow-up period was reduced to two years. Further, cases that presented as late as 2008 were accepted into the study. Therefore, the maximum follow-up for these individuals was not more than two years. Because a significant portion of the data was limited to two years of follow-up, it would be inappropriate to draw conclusions about longer survival periods. The alternative, excluding those cases with only two years of follow-up would bias the sample, selectively excluding less severe cases. Further, sample size would be greatly reduced. Such is a common limitation of retrospective studies in this disease.

9.4.5.2 Cut-points were used for MRI-based measurements

As a result of the data distribution, continuous variable categorization was considered for the three MRI-based variables. In Chapter 4, continuous data categorization was discussed as a poor practice that leads to loss of information. When consulted, statisticians agreed that, in some cases, categorization may be necessary to overcome undesirable data characteristics (Personal correspondence, Oxford Department of Statistics Consulting Service). What this work did not do is examine multiple cut-point variables, looking for an "optimum value". Specific values for categorization were predefined in the methods.

While not ideal, this practice is not prone to the p-value optimism seen with multiple testing. Despite this, we can still expect the cut-point selected to be sample dependent.

9.4.5.3 Models are basic and would benefit from refinement

The multivariate models that were constructed in this thesis are mathematically complex and require considerable computational power. However, the methodology used to construct the logistic regression, Cox proportional hazards, artificial neural network and recursive partitioning models is basic compared to that used by experts in statistics and data mining. A comprehensive approach to modeling requires the dedicated experience and input of an expert in the field.

Despite this, statistical software packages have progressed to the point that these techniques are no longer beyond the ability of the average clinical researcher. Equipped with a sound understanding of the required principles, non-experts can successfully build and interpret complex models for prediction and prognosis. While the models created in this work would benefit from refinement, one of the accomplishments of this thesis is to demonstrate that clinical researchers need not shy away from these techniques.

9.4.6 Work on extrinsic muscle position should be interpreted with caution

9.4.6.1 Distance is a sub-optimum measure of extrinsic muscle position

The hyoglossus and styloglossus are three-dimensional objects and thus, measurement of the

distance from the surface mucosa is somewhat subjective. The measurements can only be made based on the contour of the muscle in the plane of the coronal and axial slices. The "distance" measurement is also based on what is subjectively chosen as the "point of least distance" from the surface. Though five different points were chosen, the existence of a closer point cannot be excluded. A better approach would be a three-dimensional method that considers every pixel of the surface mucosa and extrinsic muscles. Such a method would objectively select the pair of points that constitutes the smallest distance.

9.4.7 The atlas is a rough approximation of extrinsic muscle position

It is worth reiterating that the tongue-atlas fusions represent only a rough approximation of extrinsic muscle invasion. A full, systematic treatment of tongue-atlas fusion requires a dedicated computer science approach to automated three-dimensional co-registration. Such an approach would also require improvement of MRI-image quality, or measures intended to improve the definition of the tongue surface contour.

9.5 Continuing the present work

9.5.1 There is a case for a prospective, multicenter study

Sample size was a major limiting factor in this work. Expansion of the sample size is only possible through the inclusion of other institutions. This is best accomplished by building upon the principles established in this work to carefully design a prospective, multicenter study. This section touches upon general considerations for such a design.

9.5.2 Prospective versus retrospective

In retrospective study design, the nature and the quality of the data is determined by the content of patient reports. Patient reports were designed with no specific research purpose in mind and, thus, data quality was generally poor. A prospective study would allow data gathering to be designed *ab initio*. It would be much easier to ensure quality and problems could be identified and corrected.

9.5.3 How many institutions?

In this work, 179 cases suitable for inclusion were identified at the JRH between 1998 and 2008, an average of 18 per year. The probability of missing records or unavailable MRI scans generally decreased in that time interval and, thus, the present rate of case identification is higher. However, the JRH has a high throughput of OCSCC cases and rates at other institutions might not be as high.

Estimating 20 valid cases per year, it is possible to deduce the number of institutions required for prospective work. If a sample size goal of 500 cases is set with 5-year follow-up, a five-institution study would require five years to gather the 500 patients and 5 years to complete follow-up. OCLNM results would be publishable at year 5, 2-year survival at year 7 and 5-year survival at year 10.

9.5.4 Which oral cavity subsites?

The estimate in the previous section is based on the inclusion of all oral cavity subsites except the lip. Inclusion of the lip would increase the available cases. However, at the moment the lip is not routinely imaged by MRI or CT and would require a modification of staging protocols.

Limiting the study to the oral tongue cases is worth considering. Measurement of volume, diameter and thickness is relatively easy due to its simple shape, well-defined appearance on MRI and the ease of tumour identification on T2-weighted pulse sequences. This removes the influence of structural heterogeneity and differences in natural history of disease among oral cavity subsites. The cost is a reduction in available cases by approximately one-half. Case availability could be increased by including both the oral and oropharyngeal tongue. However, there are issues with tumours of the base of the tongue being difficult to distinguish from the lymphoid tissue of the lingual tonsil on MRI.

9.5.5 Concerning volume of invasion

9.5.5.1 Validation and interoperator variability

MRI-based volume requires validation. Pathologists agree that this is a challenging task, requiring specialized skills and equipment (Personal correspondence Prof. K. Gatter, Consultant Pathologist, and Dr. K. Shah, Consultant Pathologist). This was beyond the scope of the present work but should

be considered prior to prospective study. Interobserver variability should also be assessed. Agreement between observers is essential to ensure that work carried out at different institutions is comparable.

9.5.5.2 Automation

Manual segmentation of the tumour volume is time consuming and is likely to be operator dependent. These problems could be remedied by an automated or semi-automated technique. Chong et al. used a seed-growing method to semi-automatically segment the tumour volume in 16 cases of oral tongue SCC [240]. The group also demonstrated that semi-automation improves interoperator variability. Fully automated techniques are dependent on image resolution. The poor resolution of MRI prevents the use of full automation at present, but it remains a possibility for future application.

9.5.6 Database design and data gathering

The database design used in this work is useful as a template for a prospective study. A redesign of the graphical user interface (GUI) would simplify data input.

A great deal of work is required to properly gather the required data. Two phases can be envisioned; primary data gathering from the patient and patient record, and then input into the database. Compliance is a chief concern. If the complexity of data gathering is high, compliance and thus the quality of available data will be low.

One way to improve compliance is to truncate the database to remove information that is not relevant to the prospective study. For example, in this work data was gathered about the pattern of tumour growth (exophytic vs. ulcerative) but contributed little to the final result. A second option would be to introduce intuitive data-gathering forms that prompt staff to gather the required information. The forms could then be passed to an individual in charge of ensuring data fidelity and inputting the information into the central database. Compliance is also dependent upon the data gatherer. A dedicated research staff person would raise compliance, but be a costly option.

9.5.7 Collating data from different institutions and data protection

One of the challenges of a multicenter study would be centralizing the data at the lead institution. The primary issue is adherence to the Data Protection Act.

All data would have to be pseudo-anonymized. Patient identifying information (i.e. hospital number, date of birth, name) would have to be kept in a separate database, linked to the main database by a unique identifier. Only the main database would be sent to the lead institution for analysis. In this way, loss of either database in isolation is less catastrophic.

Information transfer must also be carefully considered. Transfer by unsecured connection, external memory (i.e. USB, CD) or email is risky. Secure connections for data transfer do already exist between NHS institutions and are used for PACS image transfer (Personal communication, Mr. Dermot Dobson). These connections could be adapted for secure database and image transfer.

9.5.8 Improving resolution and standardized image acquisition

A large number of cases were excluded because of motion and susceptibility artefact. Elimination of such artefacts would improve case availability. Further, improved contrast, spatial resolution and slice thickness would improve the accuracy of MRI volume determination and MRI-based measurement. Standardizing the MRI acquisition protocol would remove a source of measurement uncertainty. A proposal for a standardized imaging protocol is shown in Table 9.2.

Table 9.2: Proposed parameters for a standardized MRI protocol.

Pulse sequences	*Coronal T1-weighted, fat-saturated FSE*
	Axial T1-weighted, fat-saturated FSE
	Coronal T2-weighted, fat-saturated FSE
	Axial T2-weighted, fat-saturated FSE
RF coil	*Head and neck*
Field of view	*20 cm X 20 cm*
Pixel matrix	*512 X 512*
Slice thickness	*4*
Slice spacing	*1*
Zoom factor	*1*
Number of excitations	*1*

9.5.9 Data analysis

The R-Project scripts produced for this thesis were designed to operate under a variety of conditions and could be adapted for analysis of data gathered by a prospective study.

9.6 Conclusions

This work presented in this thesis is an original contribution to the field of oral cavity cancer research. This thesis has determined that there is capacity for improvement in current efforts to determine the natural history and prognosis of oral cavity squamous cell carcinoma.

This thesis has shown that the depth of invasion literature, while promising, suffers from poor statistical practice. The use of continuous variable categorization and multiple testing is widespread, and contributes to cut-point variability and false-positive tests. Depth, as a predictor of OCLNM and survival, must be questioned.

This thesis has also shown that there are a number of patient and tumour related factors that are independent predictors of OCLNM and survival. MRI-based tumour volume, categorized at 3.0 cm^3, is at least equivalent to MRI-based thickness (11.0 mm) for OCLNM prediction, and may, in fact, be the superior predictor of OCLNM.

For survival modeling, artificial neural networks are superior to regression and recursive partitioning. There is a case for a unified model that uses tumour volume, N-stage and midline invasion to determine prognosis. This model can be used to determine individualized probabilities of 2-year survival.

The role of the extrinsic muscles of the tongue in TNM staging of oral carcinoma need to be reconsidered. The lateral extrinsic muscles lie close to the surface of the lateral tongue and are invaded by superficial tumours. These carcinomas do not warrant the prognosis and aggressive treatment of Stage IV disease.

As a result of this work I recommend that those involved in the treatment of oral cavity cancers use

caution when interpreting the depth of invasion literature. Careful consideration is required before proposals for inclusion in TNM staging are moved forward.

MRI-based volume is suitable for a decision rule for the elective treatment of the neck. Patients with volume less than 3.0 cm^3 have only a 5% risk of occult metastasis and may best be treated with frequent monitoring.

Volume is a superior predictor of ACS and DRS compared to diameter. Suggested cut-points for incorporation into T1-3 criteria are 0.5 cm^3 and 8.0 cm^3. Further, the oral tongue should be given its own T4a classification based on invasion of the genioglossus and extension to the floor of the mouth. Extrinsic muscle invasion should be removed from the T4a (oral cavity) category.

Further change to the TNM criteria for the oral cavity would be detrimental to retrospective and prospective research. Modifications, while initially beneficial for prognosis, would be quickly outpaced by ongoing research. Surgeons should consider model based approaches that provide individualized prognostic information. I recommend an artificial neural network model that uses tumour diameter, N-stage and midline invasion.

References

1. Pataridis, K., *3 Dimensional modeling of the muscles of the human tongue and implications to surgical management of cancer [unpublished DPhil thesis]*. 2007. p. 1-272.
2. Moore, K.L. and A.F. Dalley, *Clinically oriented anatomy*, ed. K.L. Moore and A.F. Dalley. 1999, Philadelphia ; London: Lippincott Williams & Wilkins. 1167.
3. Shah, J.P., N.W. Johnson, and J.G. Batsakis, *Oral cancer*, ed. J.P. Shah, N.W. Johnson, and J.G. Batsakis. 2003, London: Dunitz. 496.
4. Bradford, C.R., et al., *Chapter 4: Oral Cavity Procedures*, in *ACS Surgery: Principles and Practice 2006*, W.W. Souba, Editor. 2007, WebMD Professional Publishing: New York, NY.
5. Shah, J.P., *Cancer of the head and neck*. American Cancer Society. 2001, London: BC Decker. 484.
6. TV, A. *Anatomy TV*. 2006 [cited 2009 February 15, 2009]; Available from: http://www.anatomytv.com.
7. Robbins, K.T., S. Samant, and C.W. Cummings, *Chapter 116: Neck Dissection*, in *Otolaryngology: Head and Neck Surgery*, C.W. Cummings, et al., Editors. 2005, Mosby: Philadelphia, Penn.
8. Bushberg, J.T., et al., *Chapter 14: Nuclear Magnetic resonance*, in *The essential physics of medical imaging*, J.T. Bushberg, et al., Editors. 2002, Lippincott Williams & Wilkins: Philadelphia ; London. p. 373-414.
9. Hoa, D., A. Micheau, and G. Gahide. *eMRI.com: Interactive Course About MRI Physics*. 2007 [cited 2007; Available from: http://www.imaios.com/en/e-Courses/e-MRI/NMR.
10. Westbrook, C., C.K. Roth, and J. Talbot, *MRI in practice*. 3rd ed. 2005, Oxford: Blackwell Publishing. 410.
11. Higgins, D.M. *ReviseMRI.com*. 2003 [cited 2009 February 15, 2009]; Available from: http://www.revisemri.com.
12. Bushberg, J.T., et al., *Chapter 15: Magnetic Resonance Imaging*, in *The essential physics of medical imaging*, J.T. Bushberg, et al., Editors. 2002, Lippincott Williams & Wilkins: Philadelphia ; London. p. 415-468.
13. Greene, F.L., American Joint Committee on Cancer, and American Cancer Society, *AJCC cancer staging manual*. 6th ed. 2002, New York: Springer-Verlag. xiv, 421 p.
14. Ferlay, J., et al., *GLOBOCAN 2002.* IARC CancerBase, 2002.
15. Statistics, O.f.N. *Cancer Survival: England and Wales, 1991-2001, less common cancers, by age group dataset*. 2001 [cited 2009 February15, 2009]; Available from: http://www.statistics.gov.uk/statbase/ssdataset.asp?vlnk=8904.
16. Scully, C. and S. Porter, *ABC of oral health. Oral cancer.* BMJ (Clinical research ed.), 2000. **321**(7253): p. 97-100.
17. Hoffmann, D. and M.V. Djordjevic, *Chemical composition and carcinogenicity of smokeless tobacco.* Adv Dent Res, 1997. **11**(3): p. 322-329.
18. Mashberg, A., L. Garfinkel, and S. Harris, *Alcohol as a primary risk factor in oral squamous carcinoma.* CA, 1981. **31**(3): p. 146-155.
19. Keller, A.Z., *Alcohol, tobacco and age factors in the relative frequency of cancer among males with and without liver cirrhosis.* Am J Epi, 1977. **106**(3): p. 194-202.
20. Choi, S.Y. and H. Kahyo, *Effect of cigarette smoking and alcohol consumption in the aetiology of cancer of the oral cavity, pharynx and larynx.* Int J Epi, 1991. **20**(4): p. 878-885.
21. Bouda, M., et al., *"High risk" HPV types are frequently detected in potentially malignant and malignant oral lesions, but not in normal oral mucosa.* Mod Pathol, 2000. **13**(6): p. 644-53.
22. Sand, L.P., et al., *Prevalence of Epstein-Barr virus in oral squamous cell carcinoma, oral lichen planus, and normal oral mucosa.* Oral Surg Oral Med Oral Pathol Oral Radiol Endod, 2002. **93**(5): p. 586-92.

23. Gonzalez-Moles, M., et al., *Epstein-Barr virus and oral squamous cell carcinoma in patients without HIV infection: viral detection by polymerase chain reaction.* Microbios, 1998. **96**(383): p. 23-31.
24. Mao, E.J. and C.J. Smith, *Detection of Epstein-Barr virus (EBV) DNA by the polymerase chain reaction (PCR) in oral smears from healthy individuals and patients with squamous cell carcinoma.* J Oral Pathol Med, 1993. **22**(1): p. 12-7.
25. Cruz, I., et al., *No direct role for Epstein-Barr virus in oral carcinogenesis: a study at the DNA, RNA and protein levels.* Int J Cancer, 2000. **86**(3): p. 356-61.
26. Iamaroon, A., et al., *Co-expression of p53 and Ki67 and lack of EBV expression in oral squamous cell carcinoma.* J Oral Pathol Med, 2004. **33**(1): p. 30-6.
27. Nagao, Y., et al., *High prevalence of hepatitis C virus antibody and RNA in patients with oral cancer.* J Oral Pathol Med, 1995. **24**(8): p. 354-60.
28. Porter, S.R., et al., *Development of squamous cell carcinoma in hepatitis C virus-associated lichen planus.* Oral Oncol, 1997. **33**(1): p. 58-9.
29. Goldstein, A.M., et al., *Familial risk in oral and pharyngeal cancer.* Eur J Cancer B Oral Oncol, 1994. **30B**(5): p. 319-22.
30. Brown, L.M., et al., *Family cancer history and susceptibility to oral carcinoma in Puerto Rico.* Cancer, 2001. **92**(8): p. 2102-8.
31. Copper, M.P., et al., *Role of genetic factors in the etiology of squamous cell carcinoma of the head and neck.* Arch Otolaryngol Head Neck Surg, 1995. **121**(2): p. 157-60.
32. Foulkes, W.D., et al., *Familial risks of squamous cell carcinoma of the head and neck: retrospective case-control study.* BMJ, 1996. **313**(7059): p. 716-21.
33. Bongers, V., et al., *The relation between cancer incidence among relatives and the occurrence of multiple primary carcinomas following head and neck cancer.* Cancer Epidemiol Biomarkers Prev, 1996. **5**(8): p. 595-8.
34. Ankathil, R., et al., *Is oral cancer susceptibility inherited? Report of five oral cancer families.* Eur J Cancer B Oral Oncol, 1996. **32B**(1): p. 63-7.
35. La Vecchia, C., et al., *Epidemiology and prevention of oral cancer.* Oral Oncol, 1997. **33**(5): p. 302-12.
36. Tavani, A., et al., *Diet and risk of oral and pharyngeal cancer. An Italian case-control study.* Eur J Cancer Prev, 2001. **10**(2): p. 191-5.
37. Greer, R.O., *Pathology of malignant and premalignant oral epithelial lesions.* Otolaryngol Clin North Am, 2006. **39**(2): p. 249-75.
38. Kumar, V., R.S. Cotran, and S.L. Robbins, *Robbins basic pathology*. 7th ed. 2003, Philadelphia, PA: Saunders. xii, 873 p.
39. Crissman, J.D., et al., *Prognostic value of histopathologic parameters in squamous cell carcinoma of the oropharynx.* Cancer, 1984. **54**(12): p. 2995-3001.
40. Yamamoto, E., A. Miyakawa, and G. Kohama, *Mode of invasion and lymph node metastasis in squamous cell carcinoma of the oral cavity.* Head Neck Surg, 1984. **6**(5): p. 938-947.
41. Million, R.R., et al., *Oral Cavity*, in *Management of head and neck cancer : a multidisciplinary approach*, Anonymous, Editor. 1994, Lippincott: Philadelphia. p. 321-400.
42. Chong, V., *Oral cavity cancer.* Cancer Imaging, 2005. **5 Spec No A**: p. S49-52.
43. Boyle, J.O., E.W. Strong, and J.P. Shah, *Oral cavity cancer*, in *Cancer of the Head and Neck*, J.P. Shah, Editor. 2001, BC Decker: London. p. 100-125.
44. Asakage, T., et al., *Tumor thickness predicts cervical metastasis in patients with stage I/II carcinoma of the tongue.* Cancer, 1998. **82**(8): p. 1443-1448.
45. Hayashi, T., et al., *The relationship of primary tumor thickness in carcinoma of the tongue to subsequent lymph node metastasis.* Dento Maxillo Facial Radiol, 2001. **30**(5): p. 242-245.
46. O-charoenrat, P., et al., *Tumour thickness predicts cervical nodal metastases and survival in early oral tongue cancer.* Oral Oncol, 2003. **39**(4): p. 386-390.

47. Po Wing Yuen, A., et al., *Prognostic factors of clinically stage I and II oral tongue carcinoma-A comparative study of stage, thickness, shape, growth pattern, invasive front malignancy grading, Martinez-Gimeno score, and pathologic features.* Head Neck, 2002. **24**(6): p. 513-20.
48. Shah, J.P., *Patterns of cervical lymph node metastasis from squamous carcinomas of the upper aerodigestive tract.* Am J Surg, 1990. **160**(4): p. 405-409.
49. Woolgar, J.A., *Histological distribution of cervical lymph node metastases from intraoral/oropharyngeal squamous cell carcinomas.* Br J Oral Maxillofac Surg, 1999. **37**(3): p. 175-180.
50. Greenberg, J.S., et al., *Extent of extracapsular spread: a critical prognosticator in oral tongue cancer.* Cancer, 2003. **97**(6): p. 1464-1470.
51. Kotwall, C., et al., *Metastatic patterns in squamous cell cancer of the head and neck.* Am J Surg, 1987. **154**(4): p. 439-442.
52. Myers, J.N., et al., *Extracapsular spread. A significant predictor of treatment failure in patients with squamous cell carcinoma of the tongue.* Cancer, 2001. **92**(12): p. 3030-3036.
53. Cummings, C.W., *Otolaryngology--head & neck surgery*. 3rd ed. 1998, St. Louis: Mosby.
54. Goldstein, D.P., et al., *Chapter 2: Oral Cavity Lesions*, in *ACS Surgery: Principles and Practice 2006*, W.W. Souba, et al., Editors. 2006, WebMD Professional Publishing: New York, NY.
55. Beenken, S.W., et al., *Chapter 15: Head and Neck Tumors*, in *Current surgical diagnosis & treatment*, G.M. Doherty and L.W. Way, Editors. 2006, Lang/McGraw-Hill: New York; Toronto. p. 1453.
56. Ferlito, A., et al., *Neck dissection: then and now.* Auris Nasus Larynx, 2006. **33**(4): p. 365-74.
57. Grandi, C., et al., *Prognostic significance of lymphatic spread in head and neck carcinomas: therapeutic implications.* Head Neck Surg, 1985. **8**(2): p. 67-73.
58. Castelijns, J.A. and M.W. van den Brekel, *Imaging of lymphadenopathy in the neck.* Eur Radiol, 2002. **12**(4): p. 727-738.
59. Campana, J.P. and A.D. Meyers, *The surgical management of oral cancer.* Otolaryngol Clin North Am, 2006. **39**(2): p. 331-348.
60. Chow, J.M., et al., *Radiotherapy or surgery for subclinical cervical node metastases.* Arch Otolaryngol, 1989. **115**(8): p. 981-984.
61. Ahuja, A.T., et al., *Imaging of head and neck cancer*, ed. A.T. Ahuja, et al. 2003, London: Greenwich Medical Media. 222.
62. Unger, J.M., *The oral cavity and tongue: magnetic resonance imaging.* Radiology, 1985. **155**(1): p. 151-153.
63. Lufkin, R.B., et al., *Tongue and oropharynx: findings on MR imaging.* Radiology, 1986. **161**(1): p. 69-75.
64. Simon, L.L. and D. Rubinstein, *Imaging of oral cancer.* Otolaryngol Clin North Am, 2006. **39**(2): p. 307-17, vi.
65. Lufkin, R.B. and W.N. Hanafee, *Magnetic resonance imaging of the head and neck.* Invest Radiol, 1988. **23**(3): p. 162-169.
66. Sigal, R., et al., *CT and MR imaging of squamous cell carcinoma of the tongue and floor of the mouth.* Radiographics, 1996. **16**(4): p. 787-810.
67. Heissler, E., et al., *Value of magnetic resonance imaging in staging carcinomas of the oral cavity and oropharynx.* Int J Oral Max Surg, 1994. **23**(1): p. 22-27.
68. Vogl, T., et al., *MR imaging of the oropharynx and tongue: comparison of plain and Gd-DTPA studies.* J Comput Assist Tomo, 1988. **12**(3): p. 427-433.
69. Edge, S., et al., *AJCC Cancer Staging Manual*. 2009, New York, NY: Springer.
70. Quirke, P., et al., *The future of the TNM staging system in colorectal cancer: time for a debate?* Lancet Oncol, 2007. **8**(7): p. 651-7.
71. Hibbert, J., et al., *Prognostic factors in oral squamous carcinoma and their relation to clinical staging.* Clin Otolaryngol Allied Sci, 1983. **8**(3): p. 197-203.
72. Krause, C.J., J.G. Lee, and B.F. McCabe, *Carcinoma of the oral cavity.* Arch Otolaryngol, 1973. **97**(4): p. 354-358.

73. Lee, J.G. and W.B. Litton, *Occult regional metastasis: carcinoma of the oral tongue.* Laryngoscope, 1972. **82**(7): p. 1273-1281.
74. Lenz, M., et al., *Magnetic resonance tomography of the oral cavity, the oropharynx and the mouth floor: comparison with computed tomography.* RoFo, 1989. **150**(4): p. 425-433.
75. Moore, C., M.B. Flynn, and R.A. Greenberg, *Evaluation of size in prognosis of oral cancer.* Cancer, 1986. **58**(1): p. 158-162.
76. Al-Rajhi, N., et al., *Early stage carcinoma of oral tongue: prognostic factors for local control and survival.* Oral Oncol, 2000. **36**(6): p. 508-514.
77. Borges, A.M., S.S. Shrikhande, and B. Ganesh, *Surgical pathology of squamous carcinoma of the oral cavity: its impact on management.* Semin Surg Oncol, 1989. **5**(5): p. 310-317.
78. Brown, B., et al., *Prognostic factors in mobile tongue and floor of mouth carcinoma.* Cancer, 1989. **64**(6): p. 1195-1202.
79. Byers, R.M., et al., *Can we detect or predict the presence of occult nodal metastases in patients with squamous carcinoma of the oral tongue?* Head Neck, 1998. **20**(2): p. 138-144.
80. Fakih, A.R., R.S. Rao, and A.R. Patel, *Prophylactic neck dissection in squamous cell carcinoma of oral tongue: a prospective randomized study.* Semin Surg Oncol, 1989. **5**(5): p. 327-30.
81. Fukano, H., et al., *Depth of invasion as a predictive factor for cervical lymph node metastasis in tongue carcinoma.* Head Neck, 1997. **19**(3): p. 205-210.
82. Gluckman, J.L., et al., *Prognostic indicators for squamous cell carcinoma of the oral cavity: a clinicopathologic correlation.* Laryngoscope, 1997. **107**(9): p. 1239-1244.
83. Gonzalez-Moles, M.A., et al., *Importance of tumour thickness measurement in prognosis of tongue cancer.* Oral Oncol, 2002. **38**(4): p. 394-7.
84. Hosal, A.S., O.F. Unal, and A. Ayhan, *Possible prognostic value of histopathologic parameters in patients with carcinoma of the oral tongue.* Eur Arch Otorhinolaryngol, 1998. **255**(4): p. 216-219.
85. Kligerman, J., et al., *Supraomohyoid neck dissection in the treatment of T1/T2 squamous cell carcinoma of oral cavity.* Am J Surg, 1994. **168**(5): p. 391-394.
86. Kurokawa, H., et al., *Risk factors for late cervical lymph node metastases in patients with stage I or II carcinoma of the tongue.* Head Neck, 2002. **24**(8): p. 731-736.
87. Lim, S.C., et al., *Predictive markers for late cervical metastasis in stage I and II invasive squamous cell carcinoma of the oral tongue.* Clin Cancer Res, 2004. **10**(1 Pt 1): p. 166-172.
88. Matsuura, K., et al., *Treatment results of stage I and II oral tongue cancer with interstitial brachytherapy: maximum tumor thickness is prognostic of nodal metastasis.* Int J Radiat Oncol Biol Phys, 1998. **40**(3): p. 535-539.
89. Morton, R.P., et al., *Tumor thickness in early tongue cancer.* Arch Otolaryngol, 1994. **120**(7): p. 717-720.
90. Nakagawa, T., et al., *Neck node metastasis after successful brachytherapy for early stage tongue carcinoma.* Radiother Oncol, 2003. **68**(2): p. 129-135.
91. Nathanson, A., et al., *Evaluation of some prognostic factors in small squamous cell carcinoma of the mobile tongue: a multicenter study in Sweden.* Head Neck, 1989. **11**(5): p. 387-392.
92. Okamoto, M., et al., *Prediction of delayed neck metastasis in patients with stage I/II squamous cell carcinoma of the tongue.* J Oral Pathol Med, 2002. **31**(4): p. 227-233.
93. Shintani, S., et al., *The relationship of shape of tumor invasion to depth of invasion and cervical lymph node metastasis in squamous cell carcinoma of the tongue.* Oncology, 1997. **54**(6): p. 463-467.
94. Sparano, A., et al., *Multivariate predictors of occult neck metastasis in early oral tongue cancer.* Otolaryngol Head Neck Surg, 2004. **131**(4): p. 472-476.
95. Spiro, R.H., et al., *Predictive value of tumor thickness in squamous carcinoma confined to the tongue and floor of the mouth.* Am J Surg, 1986. **152**(4): p. 345-350.
96. Williams, J.K., et al., *Tumor angiogenesis as a prognostic factor in oral cavity tumors.* Am J Surg, 1994. **168**(5): p. 373-380.

97. Woolgar, J.A., *T2 carcinoma of the tongue: the histopathologist's perspective.* Br J Oral Maxillofac Surg, 1999. **37**(3): p. 187-193.
98. Woolgar, J.A. and J. Scott, *Prediction of cervical lymph node metastasis in squamous cell carcinoma of the tongue/floor of mouth.* Head Neck, 1995. **17**(6): p. 463-472.
99. Yamazaki, H., et al., *Tongue cancer treated with brachytherapy: is thickness of tongue cancer a prognostic factor for regional control?* Anticancer Res, 1998. **18**(2B): p. 1261-1265.
100. Yuen, A.P., et al., *A comparison of the prognostic significance of tumor diameter, length, width, thickness, area, volume, and clinicopathological features of oral tongue carcinoma.* Am J Clin Pathol, 2000. **180**(2): p. 139-143.
101. Preda, L., et al., *Relationship between histologic thickness of tongue carcinoma and thickness estimated from preoperative MRI.* Eur Radiol, 2006. **16**(10): p. 2242-2248.
102. Iwai, H., et al., *Magnetic resonance determination of tumor thickness as predictive factor of cervical metastasis in oral tongue carcinoma.* Laryngoscope, 2002. **112**(3): p. 457-461.
103. Lam, P., et al., *Correlating MRI and histologic tumor thickness in the assessment of oral tongue cancer.* Am J Roentgenol, 2004. **182**(3): p. 803-808.
104. Tetsumura, A., et al., *High-resolution magnetic resonance imaging of squamous cell carcinoma of the tongue: an in vitro study.* Dento Maxillo Facial Radiol, 2001. **30**(1): p. 14-21.
105. Chew, M.H., et al., *Significance of tumour volume measurements in tongue cancer: A novel role in staging.* ANZ J Surg, 2007. **77**(8): p. 632-637.
106. Werning, J.W., *Oral cancer: diagnosis, management, and rehabilitation*. 2007, New York: Thieme Medical Publishers. xiii, 354 p.
107. Shemen, L.J., et al., *Increase of tongue cancer in young men.* JAMA, 1984. **252**(14): p. 1857.
108. Depue, R.H., *Rising mortality from cancer of the tongue in young white males.* N Engl J Med, 1986. **315**(10): p. 647.
109. Schantz, S.P., R.M. Byers, and H. Goepfert, *Tobacco and cancer of the tongue in young adults.* JAMA, 1988. **259**(13): p. 1943-4.
110. Myers, J.N., et al., *Squamous cell carcinoma of the tongue in young adults: increasing incidence and factors that predict treatment outcomes.* Otolaryngol Head Neck Surg, 2000. **122**(1): p. 44-51.
111. Sarkaria, J.N. and P.M. Harari, *Oral tongue cancer in young adults less than 40 years of age: rationale for aggressive therapy.* Head Neck, 1994. **16**(2): p. 107-11.
112. Martin-Granizo, R., et al., *Squamous cell carcinoma of the oral cavity in patients younger than 40 years.* Otolaryngol Head Neck Surg, 1997. **117**(3 Pt 1): p. 268-75.
113. Lo, W.L., et al., *Outcomes of oral squamous cell carcinoma in Taiwan after surgical therapy: factors affecting survival.* J Oral Maxillofac Surg, 2003. **61**(7): p. 751-8.
114. Nguyen, T.V. and B. Yueh, *Weight loss predicts mortality after recurrent oral cavity and oropharyngeal carcinomas.* Cancer, 2002. **95**(3): p. 553-62.
115. Leite, I.C. and S. Koifman, *Survival analysis in a sample of oral cancer patients at a reference hospital in Rio de Janeiro, Brazil.* Oral Oncol, 1998. **34**(5): p. 347-52.
116. Ribeiro, K.C., L.P. Kowalski, and M.R. Latorre, *Impact of comorbidity, symptoms, and patients' characteristics on the prognosis of oral carcinomas.* Arch Otolaryngol Head Neck Surg, 2000. **126**(9): p. 1079-85.
117. Suzuki, M., et al., *Clinicopathological factors related to cervical lymph node metastasis in a patient with carcinoma of the oral floor.* Acta Otolaryngol Suppl, 2007(559): p. 129-35.
118. Ries, L., et al., *SEER Cancer Statistics Review, 1975-2003.* Bethesda, MD: National Cancer Institute, 2006.
119. Llewellyn, C.D., N.W. Johnson, and K.A. Warnakulasuriya, *Risk factors for squamous cell carcinoma of the oral cavity in young people--a comprehensive literature review.* Oral Oncol, 2001. **37**(5): p. 401-18.
120. UK, C.R. *Oral Cancer Statistics - UK*. 2005 [cited 2010 January 20, 2010]; Available from: http://info.cancerresearchuk.org/cancerstats/types/oral/index.htm.

121. Massano, J., et al., *Oral squamous cell carcinoma: review of prognostic and predictive factors.* Oral Surg Oral Med Oral Pathol Oral Radiol Endod, 2006. **102**(1): p. 67-76.
122. Mayne, S.T., et al., *Alcohol and tobacco use prediagnosis and postdiagnosis, and survival in a cohort of patients with early stage cancers of the oral cavity, pharynx, and larynx.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(12): p. 3368-74.
123. Deleyiannis, F.W., et al., *Alcoholism: independent predictor of survival in patients with head and neck cancer.* J Natl Cancer Inst, 1996. **88**(8): p. 542-9.
124. Do, K.A., et al., *Second primary tumors in patients with upper aerodigestive tract cancers: joint effects of smoking and alcohol (United States).* Cancer Causes Control, 2003. **14**(2): p. 131-8.
125. Mansour, O.I., C.H. Snyderman, and F. D'Amico, *Association between tobacco use and metastatic neck disease.* Laryngoscope, 2003. **113**(1): p. 161-6.
126. Singh, B., et al., *Validation of the Charlson comorbidity index in patients with head and neck cancer: a multi-institutional study.* Laryngoscope, 1997. **107**(11 Pt 1): p. 1469-75.
127. Sabin, S.L., et al., *The impact of comorbidity and age on survival with laryngeal cancer.* Ear Nose Throat J, 1999. **78**(8): p. 578, 581-4.
128. Datema, F.R., et al., *Impact of comorbidity on short-term mortality and overall survival of head and neck cancer patients.* Head Neck, 2009.
129. Piccirillo, J.F., *Importance of comorbidity in head and neck cancer.* Laryngoscope, 2000. **110**(4): p. 593-602.
130. Kaplan, M.H. and A.R. Feinstein, *The importance of classifying initial co-morbidity in evaluatin the outcome of diabetes mellitus.* J Chronic Dis, 1974. **27**(7-8): p. 387-404.
131. Piccirillo, J., et al., *Inclusion of comorbidity into oncology data registries.* J Registry Manage, 1999. **26**(2): p. 66-70.
132. Paleri, V. and R.G. Wight, *Applicability of the adult comorbidity evaluation - 27 and the Charlson indexes to assess comorbidity by notes extraction in a cohort of United Kingdom patients with head and neck cancer: a retrospective study.* J Laryngol Otol, 2002. **116**(3): p. 200-5.
133. Tiwari, R., *Squamous cell carcinoma of the superior gingivolabial sulcus.* Oral Oncol, 2000. **36**(5): p. 461-5.
134. Silverman, S. and American Cancer Society, *Oral cancer*. 5th ed. 2003, Hamilton, ON; Lewiston, NY: B.C. Decker. 212 p.
135. Ambrosch, P., et al., *Micrometastases in carcinoma of the upper aerodigestive tract: detection, risk of metastasizing, and prognostic value of depth of invasion.* Head Neck, 1995. **17**(6): p. 473-479.
136. Stevens, M.H., et al., *Computed tomography of cervical lymph nodes. Staging and management of head and neck cancer.* Arch Otolaryngol, 1985. **111**(11): p. 735-9.
137. Haberal, I., et al., *Which is important in the evaluation of metastatic lymph nodes in head and neck cancer: palpation, ultrasonography, or computed tomography?* Otolaryngol Head Neck Surg, 2004. **130**(2): p. 197-201.
138. Stuckensen, T., et al., *Staging of the neck in patients with oral cavity squamous cell carcinomas: a prospective comparison of PET, ultrasound, CT and MRI.* J Craniomaxillofac Surg, 2000. **28**(6): p. 319-24.
139. van den Brekel, M.W., et al., *Cervical lymph node metastasis: assessment of radiologic criteria.* Radiology, 1990. **177**(2): p. 379-84.
140. Yonetsu, K., et al., *Contribution of doppler sonography blood flow information to the diagnosis of metastatic cervical nodes in patients with head and neck cancer: assessment in relation to anatomic levels of the neck.* AJNR Am J Neuroradiol, 2001. **22**(1): p. 163-9.
141. Curtin, H.D., et al., *Comparison of CT and MR imaging in staging of neck metastases.* Radiology, 1998. **207**(1): p. 123-30.
142. King, A.D., et al., *Necrosis in metastatic neck nodes: diagnostic accuracy of CT, MR imaging, and US.* Radiology, 2004. **230**(3): p. 720-6.

143. Steinkamp, H.J., et al., *Enlarged cervical lymph nodes at helical CT.* Radiology, 1994. **191**(3): p. 795-8.
144. Silverman, S., Jr., *Demographics and occurrence of oral and pharyngeal cancers. The outcomes, the trends, the challenge.* J Am Dent Assoc, 2001. **132 Suppl**: p. 7S-11S.
145. Merino, O.R., R.D. Lindberg, and G.H. Fletcher, *An analysis of distant metastases from squamous cell carcinoma of the upper respiratory and digestive tracts.* Cancer, 1977. **40**(1): p. 145-51.
146. Greenberg, J.S., et al., *Disparity in pathologic and clinical lymph node staging in oral tongue carcinoma. Implication for therapeutic decision making.* Cancer, 2003. **98**(3): p. 508-15.
147. Woolgar, J.A., et al., *Cervical lymph node metastasis in oral cancer: the importance of even microscopic extracapsular spread.* Oral Oncol, 2003. **39**(2): p. 130-7.
148. American Joint Committee on Cancer Staging and End Results Reporting. and American Cancer Society., *Manual for staging of cancer, 1977*. 1977, Chicago: American Joint Committee for Cancer Staging and End-Results Reporting. 174 p.
149. Soo, K.C., et al., *Squamous carcinoma of the gums.* Am J Surg, 1988. **156**(4): p. 281-5.
150. Brown, J.S. and H. Lewis-Jones, *Evidence for imaging the mandible in the management of oral squamous cell carcinoma: a review.* Br J Oral Maxillofac Surg, 2001. **39**(6): p. 411-8.
151. Kowalski, L.P., et al., *Factors influencing contralateral lymph node metastasis from oral carcinoma.* Head Neck, 1999. **21**(2): p. 104-10.
152. Mallet, Y., et al., *Head and neck cancer in young people: a series of 52 SCCs of the oral tongue in patients aged 35 years or less.* Acta Otolaryngol, 2009. **129**(12): p. 1503-8.
153. Weber, R.S., et al., *Squamous cell carcinoma of the soft palate, uvula, and anterior faucial pillar.* Otolaryngol Head Neck Surg, 1988. **99**(1): p. 16-23.
154. Helliwell, T. and J. Woolgar, *Minimum dataset for histopathology reports on head and neck carcinomas and salivary neoplasms*, T.R.C.o. Pathologists, Editor. 2005.
155. Piccirillo, J.F., et al., *Development of a new head and neck cancer-specific comorbidity index.* Arch Otolaryngol Head Neck Surg, 2002. **128**(10): p. 1172-9.
156. Takes, R.P., *Staging of the neck in patients with head and neck squamous cell cancer: imaging techniques and biomarkers.* Oral Oncol, 2004. **40**(7): p. 656-67.
157. Kosunen, A., et al., *Reduced expression of hyaluronan is a strong indicator of poor survival in oral squamous cell carcinoma.* Oral Oncol, 2004. **40**(3): p. 257-63.
158. Kurokawa, H., et al., *The high prognostic value of the histologic grade at the deep invasive front of tongue squamous cell carcinoma.* J Oral Pathol Med, 2005. **34**(6): p. 329-33.
159. Mendelson, B.C., J.E. Woods, and O.H. Beahrs, *Neck dissection in the treatment of carcinoma of the anterior two-thirds of the tongue.* Surg Gynecol Obstet, 1976. **143**(1): p. 75-80.
160. Frierson, H.F., Jr. and P.H. Cooper, *Prognostic factors in squamous cell carcinoma of the lower lip.* Hum Pathol, 1986. **17**(4): p. 346-54.
161. Close, L.G., et al., *Microvascular invasion in cancer of the oral cavity and oropharynx.* Arch Otolaryngol, 1987. **113**(11): p. 1191-1195.
162. Rasgon, B.M., et al., *Relation of lymph-node metastasis to histopathologic appearance in oral cavity and oropharyngeal carcinoma: a case series and literature review.* Laryngoscope, 1989. **99**(11): p. 1103-10.
163. Rahima, B., et al., *Prognostic significance of perineural invasion in oral and oropharyngeal carcinoma.* Oral Surg Oral Med Oral Pathol Oral Radiol Endod, 2004. **97**(4): p. 423-431.
164. Shingaki, S., et al., *Evaluation of histopathologic parameters in predicting cervical lymph node metastasis of oral and oropharyngeal carcinomas.* Oral Surg Oral Med Oral Pathol, 1988. **66**(6): p. 683-8.
165. Poleksic, S. and H.J. Kalwaic, *Prognostic value of vascular invasion in squamous cell carcinoma of the head and neck.* Plast Reconstr Surg, 1978. **61**(2): p. 234-40.
166. Martinez-Gimeno, C., et al., *Squamous cell carcinoma of the oral cavity: a clinicopathologic scoring system for evaluating risk of cervical lymph node metastasis.* Laryngoscope, 1995. **105**(7 Pt 1): p. 728-33.

167. McGavran, M.H., W.C. Bauer, and J.H. Ogura, *The incidence of cervical lymph node metastases from epidermoid carcinoma of the larynx and their relationship to certain characteristics of the primary tumor. A study based on the clinical and pathological findings for 96 patients treated by primary en bloc laryngectomy and radical neck dissection.* Cancer, 1961. **14**: p. 55-66.
168. Ross, G.L., et al., *Improved staging of cervical metastases in clinically node-negative patients with head and neck squamous cell carcinoma.* Ann Surg Oncol, 2004. **11**(2): p. 213-8.
169. Clark, J.R., et al., *Established prognostic variables in N0 oral carcinoma.* Otolaryngol Head Neck Surg, 2006. **135**(5): p. 748-753.
170. Fagan, J.J., et al., *Perineural invasion in squamous cell carcinoma of the head and neck.* Arch Otolaryngol, 1998. **124**(6): p. 637-640.
171. Rahima, B., et al., *Prognostic significance of perineural invasion in oral and oropharyngeal carcinoma.* Oral Surg Oral Med Oral Pathol Oral Radiol Endod, 2004. **97**(4): p. 423-31.
172. O'Brien, C.J., et al., *Surgical treatment of early-stage carcinoma of the oral tongue--wound adjuvant treatment be beneficial?* Head Neck Surg, 1986. **8**(6): p. 401-8.
173. Soo, K.C., et al., *Prognostic implications of perineural spread in squamous carcinomas of the head and neck.* Laryngoscope, 1986. **96**(10): p. 1145-8.
174. Maddox, W.A., *Hayes Martin lecture. Vicissitudes of head and neck cancer.* Am J Surg, 1984. **148**(4): p. 428-32.
175. de Visscher, J.G., et al., *Surgical treatment of squamous cell carcinoma of the lower lip: evaluation of long-term results and prognostic factors--a retrospective analysis of 184 patients.* J Oral Maxillofac Surg, 1998. **56**(7): p. 814-20; discussion 820-1.
176. Lydiatt, D.D., et al., *Treatment of stage I and II oral tongue cancer.* Head Neck, 1993. **15**(4): p. 308-12.
177. Woolgar, J.A. and A. Triantafyllou, *A histopathological appraisal of surgical margins in oral and oropharyngeal cancer resection specimens.* Oral Oncol, 2005. **41**(10): p. 1034-43.
178. Yilmaz, T., et al., *Prognostic significance of vascular and perineural invasion in cancer of the larynx.* Am J Otolaryngol, 1998. **19**(2): p. 83-8.
179. Unal, O.F., A. Ayhan, and A.S. Hosal, *Prognostic value of p53 expression and histopathological parameters in squamous cell carcinoma of oral tongue.* J Laryngol Otol, 1999. **113**(5): p. 446-50.
180. Conte, C.C., et al., *Clinical and pathologic prognostic variables in oropharyngeal squamous cell carcinoma.* Am J Surg, 1989. **157**(6): p. 582-4.
181. Hicks, W.L., Jr., et al., *Squamous cell carcinoma of the floor of mouth: a 20-year review.* Head Neck, 1997. **19**(5): p. 400-5.
182. Loree, T.R. and E.W. Strong, *Significance of positive margins in oral cavity squamous carcinoma.* Am J Surg, 1990. **160**(4): p. 410-4.
183. Byers, R.M., et al., *The prognostic and therapeutic value of frozen section determinations in the surgical treatment of squamous carcinoma of the head and neck.* Am J Surg, 1978. **136**(4): p. 525-8.
184. Hinerman, R.W., et al., *Postoperative irradiation for squamous cell carcinoma of the oral cavity: 35-year experience.* Head Neck, 2004. **26**(11): p. 984-94.
185. Spiro, R.H., et al., *Pattern of invasion and margin assessment in patients with oral tongue cancer.* Head Neck, 1999. **21**(5): p. 408-13.
186. van Heerden, W.F.P. and A.W. van Zyl, *Surgical pathology of oral cancer.* Diagn Histopathol, 2009. **15**(6): p. 296-302.
187. Great Britain., J. Mullock, and P. Leigh-Pollitt, *The 1998 Data Protection Act*. The point of law. 1999, London: Stationery Office. vii, 192 p.
188. Health, D.o. *Data Protection Act 1998: Guidance*. [Website] 2009 [cited 2010 February 11, 2010]; Available from: http://www.dh.gov.uk/en/Managingyourorganisation/Informationpolicy/Recordsmanagement/DH_4000489.

189. Rickards, L., *Living in Britain: Results from the 2002 General Household Survey*. 2004: The Stationery Office/Tso.
190. Greene, F.L., AJCC, and ACS, *AJCC cancer staging handbook : from the AJCC cancer staging manual*. 6th ed. 2002, New York: Springer. xv, 469 p.
191. Weinberger, P.M., et al., *Use of combination proteomic analysis to demonstrate molecular similarity of head and neck squamous cell carcinoma arising from different subsites.* Arch Otolaryngol Head Neck Surg, 2009. **135**(7): p. 694-703.
192. Woolgar, J.A. and T.R. Helliwell, *Dataset for histopathology reports on head and neck carcinomas and salivary gland neoplasms*. 2nd ed, ed. J.A. Woolgar and T.R. Helliwell. 2005, London: Royal College of Pathologists.
193. Pentenero, M., S. Gandolfo, and M. Carrozzo, *Importance of tumor thickness and depth of invasion in nodal involvement and prognosis of oral squamous cell carcinoma: a review of the literature.* Head Neck, 2005. **27**(12): p. 1080-1091.
194. Mohit-Tabatabai, M.A., et al., *Relation of thickness of floor of mouth stage I and II cancers to regional metastasis.* Am J Surg, 1986. **152**(4): p. 351-3.
195. Al-Rajhi, N., et al., *Early stage carcinoma of oral tongue: prognostic factors for local control and survival.* Oral Oncol, 2000. **36**(6): p. 508-14.
196. Moore, C., J.G. Kuhns, and R.A. Greenberg, *Thickness as prognostic aid in upper aerodigestive tract cancer.* Arch Surg, 1986. **121**(12): p. 1410-4.
197. Altman, D.G., et al., *Dangers of using "optimal" cutpoints in the evaluation of prognostic factors.* J Natl Cancer Inst, 1994. **86**(11): p. 829-35.
198. Royston, P., D.G. Altman, and W. Sauerbrei, *Dichotomizing continuous predictors in multiple regression: a bad idea.* Stat Med, 2006. **25**(1): p. 127-41.
199. MacCallum, R.C., et al., *On the practice of dichotomization of quantitative variables.* Psychol Methods, 2002. **7**(1): p. 19-40.
200. Lagakos, S.W., *Effects of mismodelling and mismeasuring explanatory variables on tests of their association with a response variable.* Stat Med, 1988. **7**(1-2): p. 257-74.
201. Hollander, N. and M. Schumacher, *On the problem of using 'optimal' cutpoints in the assessment of quantitative prognostic factors.* Onkologie, 2001. **24**(2): p. 194-9.
202. Williams, B., et al., *Finding optimal cutpoints for continuous covariates with binary and time-to-event outcomes*. 2006, Technical Reports Series.
203. Miller, R. and D. Siegmund, *Maximally selected chi square statistics.* Biometrics, 1982: p. 1011-1016.
204. Lausen, B. and M. Schumacher, *Maximally selected rank statistics.* Biometrics, 1992: p. 73-85.
205. *Cancer Survival: England and Wales: 1991-2001*. 2001 [cited 2009 February15, 2009]; Available from: http://www.statistics.gov.uk/statbase/ssdataset.asp?vlnk=8904.
206. Dawson, B. and R.G. Trapp, *Basic & clinical biostatistics*. 4th ed. 2004, New York: Lange Medical Books/McGraw-Hill, Medical Pub. Division. x, 438 p.
207. Ostir, G.V. and T. Uchida, *Logistic regression: a nontechnical review.* Am J Phys Med Rehabil, 2000. **79**(6): p. 565-72.
208. Grobman, W.A. and D.M. Stamilio, *Methods of clinical prediction.* Am J Obstet Gynecol, 2006. **194**(3): p. 888-94.
209. Agatonovic-Kustrin, S. and R. Beresford, *Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research.* J Pharm Biomed Anal, 2000. **22**(5): p. 717-27.
210. Olden, J.D., J.J. Lawler, and N.L. Poff, *Machine learning methods without tears: a primer for ecologists.* Q Rev Biol, 2008. **83**(2): p. 171-93.
211. Macones, G.A., et al., *Predicting outcomes of trials of labor in women attempting vaginal birth after cesarean delivery: a comparison of multivariate methods with neural networks.* Am J Obstet Gynecol, 2001. **184**(3): p. 409-13.
212. De'ath, G. and K. Fabricius, *Classification and regression trees: a powerful yet simple technique for ecological data analysis.* Ecology, 2000. **81**(11): p. 3178-3192.

213. Ohno-Machado, L., *Modeling medical prognosis: survival analysis techniques.* Journal of biomedical informatics, 2001. **34**(6): p. 428-439.
214. Worster, A., J. Fan, and A. Ismaila, *Understanding linear and logistic regression analyses.* CJEM, 2007. **9**(2): p. 111-3.
215. Street, W. *A neural network model for prognostic prediction*. 1998: Citeseer.
216. Chi, C., W. Street, and W. Wolberg. *Application of artificial neural network-based survival analysis on two breast cancer datasets*. 2007: Citeseer.
217. Therneau, T., et al., *The RPART package.* Website: http://cran. r-project. org/web/packages/rpart/rpart. pdf, 2008.
218. LeBlanc, M. and J. Crowley, *Relative risk trees for censored survival data.* Biometrics, 1992. **48**(2): p. 411-425.
219. Harrell, F.E., Jr., K.L. Lee, and D.B. Mark, *Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors.* Stat Med, 1996. **15**(4): p. 361-87.
220. Clark, T.G., et al., *Survival analysis part IV: further concepts and methods in survival analysis.* Br J Cancer, 2003. **89**(5): p. 781-6.
221. Steyerberg, E.W., et al., *Assessing the performance of prediction models: a framework for traditional and novel measures.* Epidemiology, 2010. **21**(1): p. 128-38.
222. Greene, F.L., *Chapter 3: Lip and oral cavity*, in *AJCC cancer staging handbook : TNM classification of malignant tumors*, F.L. Greene, Editor. 2002, Springer: New York ; London. p. 35-46.
223. Abd-El-Malek, S., *Observations on the morphology of the human tongue.* J Anat, 1939. **73**(Pt 2): p. 201-210 3.
224. Miyawaki, K., *A study of the musculature of the human tongue.* Annual bulletin of the research institute of logopedics and phoniatrics, 1974. **8**: p. 23-50.
225. Kier, X.M. and K.K. Smith, *Tongues, tentacles and trunks: the biomechanics of movement in muscular hydrostats.* J Linn Soc London Zool, 1985. **83**: p. 307-324.
226. Ackerman, M., *The visible human project.* Proc IEEE, 1998. **86**(3): p. 504-511.
227. The University of Michigan Visible Human Project. *Visible Human Browser - Female Dataset*. [Anatomic Browser] 2005 [cited 2009 26/11/2009]; Available from: http://vhp.med.umich.edu/browsers/female.html#.
228. Golding, S.J., et al., *Analysis of intralingual anatomy and its variations: a pilot study.* Eur Radiol, 2003. **13**(9): p. c11.
229. Gaige, T.A., et al., *Three dimensional myoarchitecture of the human tongue determined in vivo by diffusion tensor imaging with tractography.* J Magn Reson Imaging, 2007. **26**(3): p. 654-61.
230. American Joint Committee on Cancer Staging and End Results Reporting and American Cancer Society, *Manual for staging of cancer*. 1977, Chicago: American Joint Committee for Cancer Staging and End-Results Reporting. 174 p.
231. Beahrs, O.H., M.H. Myers, and American Joint Committee on Cancer, *Manual for staging of cancer*. 2nd ed. 1983, Philadelphia: Lippincott. xvii, 250 p.
232. Beahrs, O., et al., *Manual for staging of cancer.* Am J Clin Oncol, 1988. **11**(6): p. 686.
233. Beahrs, O.H., American Joint Committee on Cancer, and American Cancer Society, *Manual for staging of cancer*. 4th ed. 1992, Philadelphia: Lippincott. xv, 280 p.
234. Fleming, I.D., et al., *AJCC cancer staging manual*. 5th ed. 1997, Philadelphia: Lippincott-Raven. xv, 294 p.
235. Spitzer, V.M. and D.G. Whitlock, *The Visible Human Dataset: the anatomical platform for human simulation.* Anat Rec, 1998. **253**(2): p. 49-57.
236. Larsson, S.G., A. Mancuso, and W. Hanafee, *Computed tomography of the tongue and floor of the mouth.* Radiology, 1982. **143**(2): p. 493-500.

237. Muraki, A.S., et al., *CT of the oropharynx, tongue base, and floor of the mouth: normal anatomy and range of variations, and applications in staging carcinoma.* Radiology, 1983. **148**(3): p. 725-31.
238. Takemoto, H., *Morphological analyses of the human tongue musculature for three-dimensional modeling.* J Speech Lang Hear Res, 2001. **44**(1): p. 95-107.
239. Abrosoft. *Fantamorph Software*. 2009 [cited 2010 March 24, 2010]; Available from: http://www.fantamorph.com/.
240. Chong, V.F., et al., *Tongue carcinoma: tumor volume measurement.* Int J Radiat Oncol Biol Phys, 2004. **59**(1): p. 59-66.
241. Steinhart, H. and O. Kleinsasser, *Growth and spread of squamous cell carcinoma of the floor of the mouth.* Eur Arch Otorhinolaryngol, 1993. **250**(6): p. 358-61.
242. Stein, A.L. and S.R. Tahan, *Histologic correlates of metastasis in primary invasive squamous cell carcinoma of the lip.* J Cutan Pathol, 1994. **21**(1): p. 16-21.
243. Giacomarra, V., et al., *Predictive factors of nodal metastases in oral cavity and oropharynx carcinomas.* Laryngoscope, 1999. **109**(5): p. 795-9.
244. Onercl, M., T. Yilmaz, and G. Gedikoglu, *Tumor thickness as a predictor of cervical lymph node metastasis in squamous cell carcinoma of the lower lip.* Otolaryngol Head Neck Surg, 2000. **122**(1): p. 139-42.
245. Rodolico, V., et al., *Lymph node metastasis in lower lip squamous cell carcinoma in relation to tumour size, histologic variables and p27Kip1 protein expression.* Oral Oncol, 2004. **40**(1): p. 92-8.
246. Kurokawa, H., et al., *The relationship of the histologic grade at the deep invasive front and the expression of Ki-67 antigen and p53 protein in oral squamous cell carcinoma.* J Oral Pathol Med, 2005. **34**(10): p. 602-7.
247. Rodolico, V., et al., *Overexpression of cyclin D1 and interaction between p27Kip1 and tumour thickness predict lymph node metastases occurrence in lower lip squamous cell carcinoma.* Oral Oncol, 2005. **41**(3): p. 268-75.
248. Veness, M.J., et al., *Anterior tongue cancer and the incidence of cervical lymph node metastases with increasing tumour thickness: should elective treatment to the neck be standard practice in all patients?* ANZ J Surg, 2005. **75**(3): p. 101-5.
249. Jing, J., et al., *Prognostic predictors of squamous cell carcinoma of the buccal mucosa with negative surgical margins.* J Oral Maxillofac Surg, 2006. **64**(6): p. 896-901.
250. Kawano, K. and S. Yanagisawa, *Predictive value of laminin-5 and membrane type 1-matrix metalloproteinase expression for cervical lymph node metastasis in T1 and T2 squamous cell carcinomas of the tongue and floor of the mouth.* Head Neck, 2006. **28**(6): p. 525-33.
251. Kim, S.H., et al., *Correlations of oral tongue cancer invasion with matrix metalloproteinases (MMPs) and vascular endothelial growth factor (VEGF) expression.* J Surg Oncol, 2006. **93**(4): p. 330-7.
252. Wallwork, B.D., S.R. Anderson, and W.B. Coman, *Squamous cell carcinoma of the floor of the mouth: tumour thickness and the rate of cervical metastasis.* ANZ J Surg, 2007. **77**(9): p. 761-4.
253. Shiga, K., et al., *Management of the patients with early stage oral tongue cancers.* Tohoku J Exp Med, 2007. **212**(4): p. 389-96.
254. Chen, Y.W., et al., *Histopathological factors affecting nodal metastasis in tongue cancer: analysis of 94 patients in Taiwan.* Int J Oral Maxillofac Surg, 2008. **37**(10): p. 912-6.
255. Natori, T., et al., *Usefulness of intra-oral ultrasonography to predict neck metastasis in patients with tongue carcinoma.* Oral Dis, 2008. **14**(7): p. 591-9.
256. Okura, M., et al., *Tumor thickness and paralingual distance of coronal MR imaging predicts cervical node metastases in oral tongue carcinoma.* AJNR Am J Neuroradiol, 2008. **29**(1): p. 45-50.
257. Fakih, A.R., et al., *Elective versus therapeutic neck dissection in early carcinoma of the oral tongue.* American Journal of Surgery, 1989. **158**(4): p. 309-313.

258. Kim, H.C., J. Kusukawa, and T. Kameyama, *Clinicopathologic parameters in predicting cervical nodal metastasis in early squamous cell carcinoma of the oral cavity.* Kurume Med J, 1993. **40**(4): p. 183-92.
259. Sheahan, P., et al., *Effect of tumour thickness and other factors on the risk of regional disease and treatment of the N0 neck in early oral squamous carcinoma.* Clin Otolaryngol Allied Sci, 2003. **28**(5): p. 461-71.
260. Ross, G.L., et al., *Improved staging of cervical metastases in clinically node-negative patients with head and neck squamous cell carcinoma.* Ann Surg Oncol, 2004. **11**(2): p. 213-218.
261. Kane, S.V., et al., *Depth of invasion is the most significant histological predictor of subclinical cervical lymph node metastasis in early squamous carcinomas of the oral cavity.* European Journal of Surgical Oncology, 2006. **32**(7): p. 795-803.
262. Liao, C.T., et al., *Good tumor control and survivals of squamous cell carcinoma of buccal mucosa treated with radical surgery with or without neck dissection in Taiwan.* Oral Oncol, 2006. **42**(8): p. 800-9.
263. Keski-Santti, H., et al., *Predictive value of histopathologic parameters in early squamous cell carcinoma of oral tongue.* Oral Oncol, 2007. **43**(10): p. 1007-13.
264. Menezes, M.B., C.N. Lehn, and A.J. Goncalves, *Epidemiological and histopathological data and E-cadherin-like prognostic factors in early carcinomas of the tongue and floor of mouth.* Oral Oncol, 2007. **43**(7): p. 656-61.
265. Warburton, G., et al., *Histopathological and lymphangiogenic parameters in relation to lymph node metastasis in early stage oral squamous cell carcinoma.* J Oral Maxillofac Surg, 2007. **65**(3): p. 475-84.
266. Alkureishi, L.W., et al., *Does tumor depth affect nodal upstaging in squamous cell carcinoma of the head and neck?* Laryngoscope, 2008. **118**(4): p. 629-34.
267. Jin, W.L., et al., *Occult cervical lymph node metastases in 100 consecutive patients with cN0 tongue cancer.* Chin Med J (Engl), 2008. **121**(19): p. 1871-4.
268. Urist, M.M., et al., *Squamous cell carcinoma of the buccal mucosa: analysis of prognostic factors.* Am J Surg, 1987. **154**(4): p. 411-4.
269. Liao, C.T., et al., *Survival in squamous cell carcinoma of the oral cavity: differences between pT4 N0 and other stage IVA categories.* Cancer, 2007. **110**(3): p. 564-71.
270. Liao, C.T., et al., *Analysis of risk factors of predictive local tumor control in oral cavity cancer.* Ann Surg Oncol, 2008. **15**(3): p. 915-22.
271. Kuriakose, M.A., et al., *Tumour volume estimated by computed tomography as a predictive factor in carcinoma of the tongue.* Br J Oral Maxillofac Surg, 2000. **38**(5): p. 460-5.
272. Jones, K.R., et al., *Prognostic factors in the recurrence of stage I and II squamous cell cancer of the oral cavity.* Arch Otolaryngol Head Neck Surg, 1992. **118**(5): p. 483-5.
273. Kuriakose, M.A., et al., *Interleukin-12 delivered by biodegradable microspheres promotes the antitumor activity of human peripheral blood lymphocytes in a human head and neck tumor xenograft/SCID mouse model.* Head Neck, 2000. **22**(1): p. 57-63.
274. Chuang, H.C., et al., *High expression of CD105 as a prognostic predictor of early tongue cancer.* Laryngoscope, 2006. **116**(7): p. 1175-9.
275. Chien, C.Y., et al., *Clinical significance of osteopontin expression in T1 and T2 tongue cancers.* Head Neck, 2008. **30**(6): p. 776-81.

Appendix I: Adult Comorbidity Evaluation (ACE) 27

Using the following index, identify and grade the important medical comorbidities. The overall comorbidity score is defined according to the highest ranked single ailment, except in the case where two or more Grade 2 ailments occur in different organ systems. In this situation, the overall comorbidity score should be designated Grade 3.

Adapted from Paleri and Wright [132].

Cogent comorbid ailment	Grade 3 Severe decompression	Grade 2 Moderate decompensation	Grade 1 Mild decompensation
Cardiovascular system			
Myocardial Infarct	MI < 6 months	MI > 6 months ago	Old MI by ECG only, age undetermined
Angina/Coronary artery disease	Unstable angina	Chronic exertional angina Recent (< 6 months) coronary artery bypass graft (CABG) or percutaneous transluminal coronary angioplasty (PTCA) Recent (< 6 months) coronary stent	ECG or stress test evidence or catheterization evidence of coronary disease without symptoms Angina pectoris not requiring hospitalization CABG or PTCA (> 6 months) Coronary stent (> 6 months)
Congestive heart failure (CHF)	Hospitalization for CHF within past 6 months Ejection fraction < 20%	Hospitalization for CHF > 6 months prior CHF with dyspnoea which limits activities	CHF with dysponoea which has responded to treatment Exertional dyspnoea Paroxysmal nocturnal dyspnoea (PND)
Arrhythmias	Ventricular arrhythmia < 6 months	Ventricular arrhythmia > 6 months ago Chronic atrial fibrillation or flutter Pacemaker	Sick sinus syndrome
Hypertension	DBP > 130 mmHg Severe malignant papilloedema or other eye changes Encephalopathy	DBP 115–129 mmHg Secondary cardiovascular symptoms: vertigo, epistaxis, headaches	DBP 90–114 mmHg DBP < 90 mmHg while taking antihypertensive medications
Venous disease	Recent PE (< 6 months) Use of venous filter for PE's	DVT controlled with intravenous or oral anti-coagulants Old PE > 6 months	Old DVT no longer treated with intravenous or oral anti-coagulants
Peripheral arterial disease	Bypass or amputation for gangrene or arterial insufficiency < 6 months ago Untreated thoracic or abdominal aneurysm (> 6 cm)	Bypass or amputation for gangrene or arterial insufficiency > 6 months Chronic insufficiency	Intermittent claudication Untreated thoracic or abdominal aneurysm (< 6 cm) s/p abdominal or thoracic aortic aneurysm repair
Respiratory system			
	Marked pulmonary insufficiency Restrictive lung disease or COPD with dyspnoea at rest despite treatment Chronic supplemental O2 CO2 retention (pCO2 > 6.65kpa) Baseline pO2 < 6.65kpa FEV1 (< 50%)	Restrictive lung disease or COPD (chronic bronchitis, emphysema or asthma) with dyspnoea which limits activities FEV1 (51–65%)	Restrictive lung disease or COPD (chronic bronchitis, emphysema or asthma) with dyspnoea which has responded to treatment FEV1 (66%–80%)

Gastrointestinal system			
Hepatic	Portal hypertension and/or oesophageal bleeding < 6 months (encephalopathy, ascites, jaundice with total bilirubin > 34mmol/L) H/o Transplant < 6 months or acute rejection	Chronic hepatitis, cirrhosis, portal hypertension with moderate symptoms 'compensated hepatic failure'	Chronic hepatitis or cirrhosis without portal hypertension Chronic liver disease manifested on biopsy or persistently elevated bilirubin (> 51m mol/L)
Stomach/Intestine	Recent ulcers < 6 months requiring > 6 units of blood transfusion	Ulcers requiring surgery or transfusion of < 6 units of blood	Diagnosis of ulcers treated with meds Chronic malabsorption syndrome Inflammatory bowel disease (IBD) on meds or h/o with complications and/or surgery
Pancreas	Acute or chronic pancreatitis with major complications (phlegmon, abscess or pseudocyst)	Uncomplicated acute pancreatitis Chronic pancreatitis with minor complications (malabsorption, impaired glucose tolerance or GI bleeding)	Chronic pancreatitis w/o complications
Renal system			
End-stage renal disease	Creatinine > 265m mol/L with multiorgan failure, shock or sepsis Acute transplant rejection Acute dialysis	Chronic renal insufficiency with creatinine > 265mmol/L Stable transplant < 6 months Chronic dialysis	Chronic renal insufficiency with creatinine 177–265mmol/L Stable transplant > 6 months ago
Endocrine system (Code the comorbid ailments with (*) in both the Endocrine system and other organ systems if applicable)			
Diabetes mellitus	Hospitalization < 6 months for DKA Diabetes causing end-organ failure including retinopathy, neuropathy, nephropathy*, coronary disease* or peripheral arterial disease*	IDDM without complications Poorly controlled AODM	AODM controlled by oral agents only
Neurological system			
Stroke	Acute stroke with significant neurologic deficit	Old stroke with significant neurological residual	Stroke with no residual Past or recent TIA
Dementia	Severe dementia requiring full support for activities of daily living	Moderate dementia (not completely self-sufficient, needs supervising)	Mild dementia (can take care of self)
Paralysis	Paraplegia or hemiplegia requiring full support for activities of daily living	Paraplegia or hemiplegia requiring wheelchair, but able to do some self care	Paraplegia or hemiplegia but ambulatory and providing most of self care

Neuromuscular	MS, Parkinson's, myasthenia gravis or other chronic neuromuscular disorder and requiring full support for activities of daily living	MS, Parkinson's, myasthenia gravis or other chronic neuromuscular disorder, but able to do some self care	MS, Parkinson's, myasthenia gravis or other chronic neuromuscular disorder, but ambulatory and providing most of self-care
Psychiatric			
	Recent suicidal attempt Active schizophrenia	Major depression or bipolar disorder uncontrolled Schizophrenia controlled w/meds	Major depression or bipolar disorder controlled w/meds
Rheumatologic (including rheumatoid arthritis, systemic lupus, mixed connective tissue disorder, polymyositis, rheumatic polymyositis)			
	Connective tissue disorder with secondary end-organ failure (renal, cardiac, CNS)	Connective tissue disorder on steroids or immunosuppressant medications	Connective tissue disorder on NSAIDS or no treatment
Immunological system			
AIDS	Fulminant AIDS w/KS, MAI, PCP (AIDS-defining illness)	HIV+ with h/o defining illness CD4+ <200/m L	Asymptomatic HIV+ patient HIV+ with h/o AIDS defining illness CD4– > 200/m L
Malignancy (excluding cutaneous basal cell Ca., cutaneous SCCA, carcinoma in-situ and intraepithelial neoplasm)			
Solid tumour including melanoma	Uncontrolled cancer Newly diagnosed but not yet treated Metastatic solid tumour	Any controlled solid tumour without documented metastases, but initially diagnosed and treated within the last 5 years	Any controlled solid tumour without documented metastases, but initially diagnosed and treated > 5 years ago
Leukaemia or myeloma	Relapse Disease out of control	1st remission of new dx < 1 year Chronic suppressive therapy	H/o leukaemia or myeloma with last Rx > 1 year prior
Lymphoma	Relapse	1st remission of new dx < 1 year Chronic suppressive therapy	H/o lymphoma with last Rx > 1 year prior
Substance abuse			
Alcohol	Delirium tremens	Active alcohol abuse with social behavioural or medical complications	H/o alcohol abuse but not presently drinking
Illicit Drugs	Acute withdrawal syndrome	Active substance abuse with social behavioural or medical complications	H/o substance abuse but not presently using
Body weight			
Obesity		Morbid (i.e., BMI > 38)	

Appendix II: MRI scan parameters

The parameters of the scan specifically used to measure tumour volume and thickness. Cases designated as consistent with a T1 lesion, and subsequently given a measurement value of 0, were assessed on the basis of all available scans. T1 lesion scan parameters are therefore not given and represented in the table by "-".

FSE: fast spin echo; FATSAT: fat saturation; NEX: number of excitations; FOV: field of view; ET: echo train length; TR: time to repetition; TE: time to echo; STIR: short tau inversion recovery; GAD: gadolinium contrast; SE: spin echo

Case	Pulse Sequence	NEX	Axis	FOV	Matrix	Spacing	Thickness	ET	TR	TE
1	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	4640	73.3
2	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	4640	70.3
3	T2 FSE FATSAT	1	CORONAL	28 X 28	512 X 512	1	5	12	3600	68.7
4	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	5860	69.2
5	-	-	-	-	-	-	-	-	-	-
6	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	3	12	5120	69.2
7	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	0.5	6	12	4640	73.3
8	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	5120	69.7
9	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	10	4320	75.9
10	STIR	1	CORONAL	24 X 24	512 X 512	1	4	8	5680	42.1
11	STIR	1	CORONAL	26 X 26	512 X 512	0	3	7	2729	15
12	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	5960	69.2
13	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	6160	69.2
14	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	6100	69.2
15	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4520	70
16	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	4860	73.3
17	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	4400	73.3
18	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	4	12	4660	70.3
19	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	3	12	5160	70
20	COR T2 FSE	1	CORONAL	20 X 20	512 X 512	1	6	10	4000	72.7
21	-	-	-	-	-	-	-	-	-	-
22	T2 FSE FATSAT	1	AXIAL	32 X 32	512 X 512	0	5	14	5440	70.7
23	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	5	12	4300	73.3
24	STIR	1	CORONAL	26 X 26	512 X 512	1	3.5	4	3271	15
25	T1 SE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	0	720	22
26	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4080	70
27	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	5260	75.9
28	-	-	-	-	-	-	-	-	-	-
29	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	5120	69.2
30	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	4500	73.3
31	T2 FSE FATSAT	1	CORONAL	26 X 26	512 X 512	1	4	12	6000	70.2
32	-	-	-	-	-	-	-	-	-	-
33	STIR	1	AXIAL	28 X 28	256 X 256	1	4	8	5100	30.9
34	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	6	12	4820	69.2
35	COR T1 SE	1	CORONAL	20 X 20	512 X 512	0.5	5	0	520	22
36	T2 FSE FATSAT	1	CORONAL	28 X 28	512 X 512	1.5	4	12	4240	68.7
37	COR T2	1	CORONAL	32 X 32	256 X 256	1.5	5	0	4360	104.3
38	T1 FSE FATSAT GAD	1	CORONAL	22 X 22	512 X 512	1	5	4	600	15.2
39	T2 FSE FATSAT	1	CORONAL	25 X 25	512 X 512	1	4	12	4900	69.3
40	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	5	10	4300	75.9
41	T2 FSE FATSAT	1	CORONAL	28 X 28	512 X 512	1	4	12	5500	68.7
42	-	-	-	-	-	-	-	-	-	-
43	T2 FSE FATSAT	1	AXIAL	26 X 26	512 X 512	0	5	14	5040	71.7
44	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	4	12	5740	75.9
45	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	0.5	3	12	6100	69.5

Case	Pulse Sequence	NEX	Axis	FOV	Matrix	Spacing	Thickness	ET	TR	TE
46	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4080	70
47	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	6100	69.2
48	T2 FSE	1	AXIAL	26 X 26	512 X 512	0	3	14	5000	71.7
49	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	6420	69.2
50	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	5	3.5	12	4240	71.3
51	T2 FSE FATSAT	1	CORONAL	30 X 30	512 X 512	1	4	12	6800	68.1
52	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	0	5	12	4640	75.9
53	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	5900	69.2
54	T2 FSE FATSAT	1	AXIAL	26 X 26	512 X 512	0	5	12	4840	71.7
55	STIR	1	AXIAL	23 X 23	512 X 512	1	4	6	2000	15
56	-	-	-	-	-	-	-	-	-	-
57	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	7400	68.9
58	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4060	69.5
59	-	-	-	-	-	-	-	-	-	-
60	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	6100	69.2
61	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	3	12	5200	70.3
62	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	6100	69.7
63	T2 FSE FATSAT	1	CORONAL	26 X 26	512 X 512	1	6	12	5200	70.6
64	COR T2 FSE	1	CORONAL	20 X 20	512 X 512	0	5	12	4300	70.3
65	STIR	1	CORONAL	28 X 28	512 X 512	1	3.5	4	3029	15
66	T2 FSE FATSAT	1	AXIAL	48 X 48	512 X 512	2	6	20	6540	94.7
67	-	-	-	-	-	-	-	-	-	-
68	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	6620	69.2
69	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	4720	70.5
70	T2 FSE FATSAT	1	CORONAL	28 X 28	512 X 512	1	3	12	5080	68.7
71	T2 FSE FATSAT	1	CORONAL	28 X 28	512 X 512	1	4	12	5080	68.7
72	T2 SE	1	CORONAL	24 X 24	512 X 512	1.5	5	0	4540	102.2
73	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	5960	69.2
74	-	-	-	-	-	-	-	-	-	-
75	-	-	-	-	-	-	-	-	-	-
76	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	4840	69.2
77	T2 FSE FATSAT	1	AXIAL	26 X 26	512 X 512	0	5	14	5840	70.9
78	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	6	12	5500	73.3
79	STIR	1	CORONAL	26 X 26	512 X 512	1	3	5	3859	15
80	T2 FSE FATSAT	1	CORONAL	28 X 28	512 X 512	1	6	12	4900	70
81	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4080	69.8
82	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4080	70
83	-	-	-	-	-	-	-	-	-	-
84	T2 FSE FATSAT	1	CORONAL	28 X 28	512 X 512	1	5	12	4880	65.2
85	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4080	70
86	-	-	-	-	-	-	-	-	-	-
87	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	4840	69.2
88	T2 FSE	1	CORONAL	24 X 24	512 X 512	2.5	5	16	7000	92
89	T2 FSE FATSAT	1	CORONAL	26 X 26	512 X 512	1	4	12	6000	68.8
90	-	-	-	-	-	-	-	-	-	-

Case	Pulse Sequence	NEX	Axis	FOV	Matrix	Spacing	Thickness	ET	TR	TE
91	-	-	-	-	-	-	-	-	-	-
92	-	-	-	-	-	-	-	-	-	-
93	-	-	-	-	-	-	-	-	-	-
94	-	-	-	-	-	-	-	-	-	-
95	T1 FSE	1	AXIAL	30 X 30	512 X 512	0	5	4	400	16.1
96	T2 FSE	1	CORONAL	20 X 20	512 X 512	1	5	12	4020	109.9
97	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	0	5	12	5420	73.3
98	T2 FSE FATSAT	1	AXIAL	26 X 26	512 X 512	5	3.5	14	5000	72
99	STIR	1	AXIAL	24 X 24	256 X 256	1	5	0	3700	52.2
100	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4080	70
101	T2 FSE FATSAT	1	CORONAL	26 X 26	512 X 512	1	4	12	4300	68.8
102	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	6100	69.2
103	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4080	70
104	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1.5	5	12	6100	69.2
105	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	6100	69.2
106	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	6100	69.2
107	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	6	12	4800	72.3
108	STIR	1	AXIAL	26 X 26	256 X 256	1	5	0	3280	52.2
109	T1 SE	1	CORONAL	28 X 28	512 X 512	1	5	1	440	18
110	-	-	-	-	-	-	-	-	-	-
111	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4080	70
112	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	5800	70
113	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	5120	69.2
114	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	7500	69.2
115	T2 FSE FATSAT	1	CORONAL	25 X 25	512 X 512	1	4	12	5980	69.3
116	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	4	12	6200	70.3
117	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	4300	73.3
118	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4080	70
119	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	6100	69.7
120	-	-	-	-	-	-	-	-	-	-
121	T1 SE	1	CORONAL	26 X 26	256 X 256	1	6	1	600	14
122	STIR	1	AXIAL	30 X 30	512 X 512	0	5	12	4340	20.4
123	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	6	10	4640	67.7
124	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	5300	75.9
125	-	-	-	-	-	-	-	-	-	-
126	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4520	70
127	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	4300	73.3
128	-	-	-	-	-	-	-	-	-	-
129	T2 SE FATSAT	1	AXIAL	24 X 24	256 X 256	1	5	0	6680	104.3
130	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	4840	69.2
131	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	4	12	6200	73.3
132	-	-	-	-	-	-	-	-	-	-
133	T2 FSE	1	CORONAL	20 X 20	512 X 512	1	4	12	5200	109.9
134	T2 FSE FATSAT	1	AXIAL	26 X 26	512 X 512	0	5	14	5040	71.7
135	-	-	-	-	-	-	-	-	-	-

Case	Pulse Sequence	NEX	Axis	FOV	Matrix	Spacing	Thickness	ET	TR	TE
136	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4060	73.7
137	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	10	4200	75.9
138	T1 SE FATSAT GAD	1	CORONAL	28 X 28	256 X 256	1	5	1	580	14
139	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	7440	68.9
140	-	-	-	-	-	-	-	-	-	-
141	STIR	1	CORONAL	26 X 26	512 X 512	1	3	4	3512	15
142	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	4840	73.3
143	T2 FSE FATSAT	1	CORONAL	24 X 24	512 X 512	1	4	12	4580	69.2
144	T2 FSE FATSAT	1	AXIAL	27 X 27	512 X 512	0	5	14	5640	73.3
145	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	5	12	5200	73.3
146	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	3680	73.7
147	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	440	73.3
148	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	6400	69.7
149	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	4300	73.3
150	-	-	-	-	-	-	-	-	-	-
151	T2 SE	1	AXIAL	30 X 30	256 X 256	1	5	0	4000	104.3
152	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	4	12	4920	70.3
153	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	0.5	4	12	6140	69.7
154	T1 FSE	1	CORONAL	20 X 20	256 X 256	1	6	0	520	14
155	T2 FSE FATSAT	1	CORONAL	28 X 28	512 X 512	1.5	4	12	4240	68.7
156	-	-	-	-	-	-	-	-	-	-
157	T1 SE	1	CORONAL	22 X 22	512 X 512	0.5	5	1	480	7.7
158	STIR	1	AXIAL	24 X 24	512 X 512	1	5	11	8390	84
159	STIR	1	CORONAL	24 X 24	512 X 512	1	5	11	7460	83
160	T1 SE FATSAT GAD	1	AXIAL	20X17.5	512 X 448	0.5	5	1	577	12
161	-	-	-	-	-	-	-	-	-	-
162	STIR	1	CORONAL	26 X 26	448 X 448	1	5	3	537.8	18
163	-	-	-	-	-	-	-	-	-	-
164	-	-	-	-	-	-	-	-	-	-
165	-	-	-	-	-	-	-	-	-	-
166	T1 SE	1	CORONAL	20.1 X 23	512 X 448	0.5	5	1	408	8.1
167	T2 FSE	1	CORONAL	23 X 20.8	256 X 232	0.5	5	13	4780	97
168	STIR	1	AXIAL	24 X 24	512 X 512	1.2	6	11	7460	80
169	T2 FSE FATSAT	1	CORONAL	20 X 20	256 X 256	0.5	5	19	2557	125
170	STIR	1	CORONAL	24 X 24	512 X 512	1	5	11	6180	83
171	T2 FSE	1	CORONAL	20 X 20	256 X 256	0.5	5	13	4780	97
172	-	-	-	-	-	-	-	-	-	-
173	T2 FSE	1	CORONAL	24 X 24	512 X 512	0.5	5	11	2800	98
174	STIR	1	AXIAL	24 X 21.7	232 X 256	0.5	5	7	3850	70
175	T2 FSE FATSAT	1	CORONAL	30 X 30	512 X 512	0.5	5	13	6330	84
176	-	-	-	-	-	-	-	-	-	-
177	T1 SE	1	CORONAL	24 X 24	512 X 448	0.5	5	1	450	8.1
178	STIR	1	CORONAL	24 X 24	512 X 512	1	5	11	7460	83
179	T2 FSE FATSAT	1	CORONAL	26 X 26	512 X 512	1	4	12	4040	69.1
180	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4060	69.4

Case	Pulse Sequence	NEX	Axis	FOV	Matrix	Spacing	Thickness	ET	TR	TE
181	STIR	1	CORONAL	29.9 X 29.9	256 X 256	1	5	12	3560	28.1
182	-	-	-	-	-	-	-	-	-	-
183	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4080	70
184	-	-	-	-	-	-	-	-	-	-
185	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4080	70
186	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	4420	74.7
187	T2 FSE FATSAT	1	CORONAL	25.9 X 25.9	512 X 512	1	4	12	4480	69.1
188	STIR	1	CORONAL	25.9 X 25.9	512 X 512	1	3	4	3755	15
189	-	-	-	-	-	-	-	-	-	-
190	T2 FSE FATSAT	1	AXIAL	26 X 26	512 X 512	2	6	14	4820	71.7
191	-	-	-	-	-	-	-	-	-	-
192	T2 FSE FATSAT	1	CORONAL	22 X 22	512 X 512	1	4	12	4080	70
193	T1 SE FATSAT GAD	1	CORONAL	22 X 22	512 X 512	1	5	1	400	9
194	T1 FSE FATSAT GAD	1	AXIAL	26 X 26	512 X 512	0	4	2	420	17.1
195	STIR	1	CORONAL	26 X 26	512 X 512	1.7	3	4	3028	15
196	STIR	1	CORONAL	28 X 28	512 X 512	0.3	3	4	1574	15
197	T2 FSE FATSAT	1	CORONAL	20 X 20	512 X 512	1	6	12	4300	73.3
198	-	-	-	-	-	-	-	-	-	-
199	STIR	1	CORONAL	26 X 26	512 X 512	1	3	6	3948	15

Appendix IIIa: CLNM cut-points

Summary of publications with associated samples and statistical tests addressing the relationship between thickness and/or depth of invasion and cervical lymph node metastasis. The table gives the first author of the publication, the year published (Year), the oral cavity subsite studied (site), the sample size (n), the T-stages studied (T), the N-stages studied (N), the statistical test used (Test), the cut-off value used (Cut-Off), the method used to determine the cut-off value (Method) and the resulting p-value reported (p-value).

F: floor of the mouth; OC: oral cavity; OP: oropharynx; UADC: upper aerodigestive tract; T (NS): tongue not otherwise specified; OT: oral tongue; OPT: oropharyngeal tongue; VT: ventral tongue; B: buccal mucosa; Chi2: X^2 test; t: t-test; FET: Fisher's exact test; MWU: Mann-Whitney U test; Log Reg: logistic regression; cont: continuous variable (no cut-point used).

*: Depth and thickness defined as separate measurements. Thickness is listed first.

(a): Thickness measured on MRI and histopathology. MRI is listed first.

(b): Time to cervical lymph node metastasis measured on patients not undergoing neck dissection.

(c): Thickness measured on MRI.

(d) Patients treated with surgery and radiotherapy.

(e) Thickness measured on ultrasound and histopathology.

First Author	Year	Site	n	T	N	Test	Cut-Off	Method	p-Value
Frierson[160]	1986	Lip	175	T1-4	N0-3	t	Cont	Cont	<0.0001
			175	T1-4	N0-3	Chi2/FET	6mm	Undefined	<0.0001
Mohit-Tabatabai[194]	1986	F	148	T1-4	N0-3	Chi2/FET	1.5mm	Undefined	0.0001
Close[161]	1987	OC/OP	46	T2-4	N0-3	Chi2/FET	5/10mm	Undefined	0.695
Rasgon[162]	1989	UADC	22	T1-4	N0-3	Chi2/FET	5mm	Undefined	0.012
Stein[242]	1994	Lip	44	T1-4	N0-3	t	Cont	Cont	0.0001
			44	T1-4	N0-3	Chi2/FET	2mm	Median	0.028
Ambrosch[135]	1995	UADC	128	T1-4	N0-3	Chi2/FET	4mm	Undefined	<0.0001
Woolgar*[98]	1995	T (NS)/F	45	T1-4	N0-3	t/MWU	Cont	Cont	0.008
			45	T1-4	N0-3	t/MWU	Cont	Cont	0.006
Martinez-Gimeno[166]	1995	OC/OP	126	T1-4	N0-3	Chi2/FET	3/7mm	Defined	0.005
Fukano[81]	1997	OT/OPT	34	T1-4	N0-1	Chi2/FET	5mm	Multiple	0.0003
Byers[79]	1998	OT	85	T1-4	N0-3	Unknown	4/8/16mm	Undefined	0.0001
de Visscher[175]	1998	Lip	184	T1-4	N0-3	Chi2/FET	3/6mm	Undefined	<0.001
Hosal[84]	1998	OT	60	T1-4	N0-3	Log Reg	9mm	Defined	0
Giacomarra[243]	1999	OC/OP	61	T1-4	N0-3	Chi2/FET	7mm	Undefined	0.034
Gluckman[82]	1999	F/VT	41	T1	N0-3	Chi2/FET	3mm	Defined	0.51
Woolgar[97]	1999	OT	50	T2	N0-3	t	Cont	Cont	<0.0001
Yuen[100]	2000	OT	85	T1-3	N0-3	t	Cont	Cont	0.005
Onercl[244]	2002	Lip	27	T1-4	N0-3	MWU	Cont	Cont	0.0267
			27	T1-4	N0-3	Chi2/FET	5mm	Multiple	0.037
Iwai (a)[102]	2002	OT	30	T1-3	N0-2	Chi2/FET	6mm	Undefined	<0.0051
			30	T1-3	N0-2	Chi2/FET	4mm	Undefined	<0.0051
Rodolico[245]	2004	Lip	95	T1-4	N0-3	Chi2/FET	2/6mm	Undefined	<0.001
Kurokawa[246]	2005	T (NS)	124	T1-4	N0-3	Log Reg	4mm	Multiple	0.001
Rodolico (b)[247]	2005	Lip	97	T1-4	N0-3	Cox	Cont	Cont	<0.0001
			97	T1-4	N0-3	Cox	5mm	3rd Quartile	<0.0001
Veness[248]	2005	OT	99	T1-4	N0-3	Chi2/FET	5mm	Defined	0.007
Jing[249]	2006	B	45	T1-4	N0-3	t	Cont	Cont	0.022
Kawano[250]	2006	OT/F	57	T1-2	N0-3	Chi2/FET	4.5mm	Undefined	0.0073
Kim[251]	2006	OT	38	T1-2	N0-2	t	Cont	Cont	0.007
Preda *(c)[101]	2006	T (NS)	33	T1-4	N0-2	F	Cont	Cont	0.26
			24	T1-3	N0-2	F	Cont	Cont	0.57
			33	T1-4	N0-2	F	Cont	Cont	0.35
			24	T1-3	N0-2	F	Cont	Cont	0.35
Suzuki (d)[117]	2007	F	35	T1-2	N0-3	Chi2/FET	4mm	Undefined	<0.01
			48	T1-4	N0-3	Chi2/FET	4mm	?	<0.05
Wallwork[252]	2007	F	30	T1-2	N0-3	t	Cont	Cont	0.04
			30	T1-2	N0-3	Chi2/FET	7.5mm	Multiple	0.03
			53	T1-4	N0-3	t	Cont	Cont	0.004
			53	T1-4	N0-3	Chi2/FET	7.5mm	Multiple	0.001
Shiga[253]	2007	OT	28	T1-2	N0-2	Chi2/FET	3/10mm	Undefined	0.013
Chen[254]	2008	T (NS)	94	T1-4	N0-3	Chi2/FET	3mm	Undefined	0.02
Natori (e)[255]	2008	T (NS)	110	T1-4	N0-3	t	Cont	Cont	<0.0001

First Author	Year	Site	n	T	N	Test	Cut-Off	Method	p-Value
			110	T1-4	N0-3	Chi2/FET	5mm	Undefined	<0.001
			110	T1-4	N0-3	t	Cont	Cont	<0.0004
			110	T1-4	N0-3	Chi2/FET	8mm	Undefined	<0.001
Okura[256]	2008	OT	43	T1-4	N0-3	Log Reg	Cont	Cont	<0.005

Appendix IIIb: OCLNM Cut-Points

Summary of publications with associated samples and statistical tests addressing the relationship between thickness and/or depth of invasion and cervical lymph node metastasis. The table gives the first author of the publication, the year published (Year), the oral cavity subsite studied (site), the sample size (n), the T-stages studied (T), the N-stages studied (N), the statistical test used (Test), the cut-off value used (Cut-Off), the method used to determine the cut-off value (Method) and the resulting p-value reported (p-value).

F: floor of the mouth; OC: oral cavity; OP: oropharynx; UADC: upper aerodigestive tract; T (NS): tongue not otherwise specified; OT: oral tongue; OPT: oropharyngeal tongue; VT: ventral tongue; B: buccal mucosa; Chi2: X^2 test; t: t-test; FET: Fisher's exact test; MWU: Mann-Whitney U test; Log Reg: logistic regression; cont: continuous variable (no cut-point used); CA: Cochrane-Armitage Test

*: Depth and thickness defined as separate measurements. Thickness is listed first.

(a) Patients treated with surgery and chemotherapy.

(b) Patients treated with surgery, chemotherapy and radiotherapy.

First Author	Year	Site	n	T	N	Test	Cut-Off	Method	p-Value
Borges[77]	1989	OT	25	T1-4	N0	Chi2/FET	5mm	Undefined	0.007
Brown[78]	1989	OT/F	73	T1-3	N0	Chi2/FET	3/7mm	Defined	0.48
Fakih[257]	1989	OT	51	T1-2	N0	Chi2/FET	4mm	Undefined	<0.01
Kim (a)[258]	1993	OC	90	T1-2	N0	Chi2/FET	1.5/3/6mm	Undefined	<0.025
Kligerman[85]	1994	OT/F	34	T1-2	N0	Chi2/FET	4mm	Mean	0.11
Asakage[44]	1998	OT	44	T1-2	N0	Chi2/FET	4mm	Undefined	<0.001
Giacomarra[243]	1999	OC/OP	29	T1-4	N0	Chi2/FET	7mm	Undefined	0.6
Yuen[100]	2000	OT	76	T1-3	N0	Spearman	3/6/9/12mm	Defined	0.023
Kurokawa[86]	2002	T (NS)	50	T1-2	N0	Chi2/FET	4mm	Multiple	<0.01
Okamoto (b)[92]	2002	OT	59	T1-2	N0	Chi2/FET	4mm	Undefined	0.01
Yuen[47]	2002	OT	72	T1-2	N0	Chi2/FET	3/9mm	Defined	0.031
O-charoenrat[46]	2003	OT	50	T1-2	N0	Chi2/FET	5mm	Undefined	0.003
Sheahan[259]	2003	OC	63	T1-2	N0	MWU	Cont	Cont	0.0031
			63	T1-2	N0	t	Cont	Cont	0.0096
			63	T1-2	N0	Log Reg	Cont	Cont	0.0175
			63	T1-2	N0	Log Reg	5mm	Multiple	0.01
Lim[87]	2004	OT	56	T1-2	N0	Chi2/FET	4mm	Undefined	0.009
Sparano*[94]	2004	OT	45	T1-2	N0	Log Reg	Cont	Cont	0.01
			45	T1-2	N0	Chi2/FET	4mm	Undefined	0.009
			45	T1-2	N0	Log Reg	Cont	Cont	NS
			45	T1-2	N0	Chi2/FET	4mm	Undefined	0.004
Ross[260]	2004	OT/F/B	61	T1-2	N0	t	Cont	Cont	<0.05
			61	T1-2	N0	Chi2/FET	2mm	Undefined	0.063
		OT	21	T1-2	N0	Chi2/FET	2mm	Undefined	NS
Clark[169]	2006	OC	105	T1-4	N0	Chi2/FET	5mm	Multiple	0.001
Kane*[261]	2006	OC	48	T1-2	N0	Log Reg	Cont	Cont	0.046
			48	T1-2	N0	Log Reg	5mm	Defined	NS
			48	T1-2	N0	Log Reg	Cont	Cont	0.026
			48	T1-2	N0	Log Reg	5mm	Defined	0.101
Liao[262]	2006	B	27	T1	N0	Chi2/FET	6mm	Multiple	0.002
Keski-Santti[263]	2007	OT	73	T1-2	N0	MWU	Cont	Cont	<0.05
			73	T1-2	N0	Bootstrap	5.5mm	Multiple	NA
Menezes[264]	2007	OT/F	50	T1-2	N0	MWU	3/5mm	Undefined	0.029
Warburton[265]	2007	OC	29	T1-2	N0	Log Reg	Cont	Cont	0.0006
			29	T1-2	N0	Chi2/FET	2.2mm	Defined	0.003
Alkureishi[266]	2008	OC/OP	172	T1-2	N0	Chi2/FET	4mm	Multiple	<0.001
Jin[267]	2008	OT/OPT	100	T1-4	N0	CA	Cont	Cont	<0.05

Appendix IIIc: ACS Cut-Points

Summary of publications with associated samples and statistical tests addressing the relationship between thickness and/or depth of invasion and all-cause survival. The table gives the first author of the publication, the year published (Year), the oral cavity subsite studied (site), the sample size (n), the T-stages studied (T), the N-stages studied (N), the statistical test used (Test), the cut-off value used (Cut-Off), the method used to determine the cut-off value (Method) and the resulting p-value reported (p-value).

F: floor of the mouth; OC: oral cavity; T (NS): tongue not otherwise specified; OT: oral tongue; B: buccal mucosa; Chi2: X^2 test; t: t-test; FET: Fisher's exact test; Log Reg: logistic regression; Cox: Cox proportional hazards test; cont: continuous variable (no cut-point used).

*: Depth and thickness defined as separate measurements. Thickness is listed first.

(a) Patients treated with brachytherapy.

First Author	Year	Site	n	T	N	Test	Cut-Off	Method	p-Value
Spiro[95]	1986	OT/F	105	T1-3	N0	Log Rank	3/8mm	Defined	0.0031
Urist*[268]	1987	B	89	T1-4	N0-3	Log Rank	6mm	Undefined	<0.0001
			89	T1-4	N0-3	Log Rank	3mm	Undefined	<0.0001
Brown[78]	1989	OT/F	87	T1-4	N0-3	Chi2	3/7mm	Defined	0.004
Fakih[257]	1989	OT	51	T1-2	N0	Chi2/FET	4mm	Undefined	<0.01
Nathanson[91]	1989	OT	49	T1	N0	Log Rank	10mm	Undefined	<0.0164
Asakage[44]	1998	OT	44	T1-2	N0	Log Rank	4mm	Undefined	NS
Matsuura (a)[88]	1998	OT	173	T1-2	N0	Wilcoxon	8mm	Median	<0.05
Woolgar[97]	1999	OT	50	T2	N0-3	t	Cont	Cont	<0.0001
Gonzalez-Moles[83]	2002	OT	59	T1-4	N0-3	Log Rank	3mm	Defined	<0.05
Kurokawa[86]	2002	T (NS)	50	T1-2	N0	Cox	4mm	Multiple	0.016
			50	T1-2	N0	Log Rank	4mm	Multiple	0.0007
O-Charoenrat[46]	2003	OT	50	T1-2	N0	Cox	5mm	Undefined	0.0022
Lim[87]	2004	OT	56	T1-2	N0	Log Rank	4mm	Undefined	0.776
Veness[248]	2005	OT	99	T1-4	N0-3	Log Rank	5mm	Defined	0.27
Liao[269]	2007	OC	88	T1-3	N1	Log Rank	6mm	Multiple	NS
			90	T3	N0	Log Rank	8mm	Multiple	NS
			105	T4	N0	Log Rank	35mm	Multiple	0.0007
			230	T1-3	N1-2	Log Rank	25mm	Multiple	0.001
Liao[270]	2008	OC	827	T1-4	N0-3	Log Rank	10mm	Multiple	<0.0001

Appendix IIId: DRS Cut-Points

Summary of publications with associated samples and statistical tests addressing the relationship between thickness and/or depth of invasion and disease-related survival. The table gives the first author of the publication, the year published (Year), the oral cavity subsite studied (site), the sample size (n), the T-stages studied (T), the N-stages studied (N), the statistical test used (Test), the cut-off value used (Cut-Off), the method used to determine the cut-off value (Method) and the resulting p-value reported (p-value).

F: floor of the mouth; OC: oral cavity; OPT: oropharyngeal tongue; T (NS): tongue not otherwise specified; OT: oral tongue; B: buccal mucosa; Chi2: X^2 test; t: t-test; FET: Fisher's exact test; Log Reg: logistic regression; Cox: Cox proportional hazards test; cont: continuous variable (no cut-point used).

(a) Patients treated with brachytherapy.

(b) Thickness measured on CT.

First Author	Year	Site	n	T	N	Test	Cut-Off	Method	p-Value
Brown[78]	1989	OT/F	87	T1-4	N0-3	Chi2	3/7mm	Defined	0.001
Morton[89]	1994	OT	26	T1-2	N0	t	Cont	Cont	NS
			26	T1-2	N0	Chi2/FET	NA	Undefined	NS
Matsuura (a)[88]	1998	OT	173	T1-2	N0	Wilcoxon	8mm	Median	NS
Al-Rajhi[76]	2000	OT	85	T1-2	N0	Log Rank	10mm	Multiple	0.0002
Kuriakose (b)[271]	2000	OT/OPT	20	T1-4	N0-3	Log Rank	5mm	Undefined	0.05
Yuen[100]	2000	OT	85	T1-3	N0-3	t	Cont	Cont	0.049
			85	T1-3	N0-3	Spearman	3/6/9/12mm	Defined	0.042
O-charoenrat[46]	2003	OT	50	T1-2	N0	Cox	5mm	Undefined	0.0067
Clark[169]	2006	OC	105	T1-4	N0	Log Rank	5mm	Multiple	0.029
			105	T1-4	N0	Log Rank	2/8mm	Multiple	0.015
Liao[262]	2006	B	232	T1-4	N0-3	Log Reg	5mm	Multiple	0.009
Keski-Santti[263]	2007	OT	73	T1-2	N0	Log Rank	NA	Multiple	NS

Appendix IIIe: DFS Cut-Points

Summary of publications with associated samples and statistical tests addressing the relationship between thickness and/or depth of invasion and disease-free survival. The table gives the first author of the publication, the year published (Year), the oral cavity subsite studied (site), the sample size (n), the T-stages studied (T), the N-stages studied (N), the statistical test used (Test), the cut-off value used (Cut-Off), the method used to determine the cut-off value (Method) and the resulting p-value reported (p-value).

OC: oral cavity; OPT: oropharyngeal tongue; T (NS): tongue not otherwise specified; OT: oral tongue; B: buccal mucosa; Chi2: X^2 test; t: t-test; FET: Fisher's exact test; Log Reg: logistic regression; Cox: Cox proportional hazards test; cont: continuous variable (no cut-point used).

(a) Thickness measured on CT.

First Author	Year	Site	n	T	N	Test	Cut-Off	Method	p-Value
Jones[272]	1992	OC	49	T1-2	N0	Cox	Cont	Cont	0.01
Morton[89]	1994	OT	26	T1-2	N0	t	Cont	Cont	NS
			26	T1-2	N0	Chi2/FET	NA	Undefined	NS
Asakage[44]	1998	OT	44	T1-2	N0	Log Rank	4mm	Undefined	S
Al-Rajhi[76]	2000	OT	85	T1-2	N0	Log Rank	10mm	Multiple	0.0002
Kuriakose (a)[273]	2000	OT/OPT	20	T1-4	N0-3	Log Rank	5mm	Undefined	0.16
Yuen[100]	2000	OT	85	T1-3	N0-3	Log Rank	3/9mm	Defined	0.028
O-charoenrat[46]	2003	OT	50	T1-2	N0	Cox	5mm	Undefined	0.0057
Kurokawa[246]	2005	T (NS)	124	T1-4	N0-3	Log Rank	4mm	Multiple	<0.0001
Veness[248]	2005	OT	99	T1-4	N0-3	Log Rank	5mm	Defined	0.47
Liao[262]	2006	B	232	T1-4	N0-3	Log Reg	5mm	Multiple	0.005
Chuang[274]	2006	T (NS)	94	T1-2	N0-3	Log Rank	4mm	Defined	0.023
Liao[269]	2007	OC	88	T1-3	N1	Log Rank	6mm	Multiple	0.0305
			230	T1-3	N1-2	Log Rank	25mm	Multiple	0.0014
			90	T3	N0	Log Rank	8mm	Multiple	0.0022
			105	T4	N0	Log Rank	35mm	Multiple	0.0385
Liao[270]	2008	OC	827	T1-4	N0-3	Log Rank	10mm	Multiple	<0.0001
Chien[275]	2008	OT	94	T1-2	N0-3	Log Rank	4mm	Defined	0.023
			49	T1-2	N0	Cox	5mm	?	NA

Appendix IV-A: Patient Data Summary

Summary of data collect from 199 patient records. For categorical data, the number of cases in each category are given (n) as well as the percentage of the total (n%). Missing values are given as "NA". For continuous data, the mean with 95% confidence interval (CI) and the median with the interquartile range (IQR) are reported.

n: number of cases; CI: confidence interval; ACE: adult comorbidity scale; OC: oral cavity; MRI: magnetic resonance imaging; CLN: cervical lymph node; ECS: extracapsular spread; JRH: John Radcliffe Hospital; UCLH: University College London Hospital; NA: not available; PVI: perivascular invasion; PNI: perineural invasion; SND: selective neck dissection; MRND: modified radical neck dissection; BND: bilateral neck dissection; NOS: not otherwise specified; OCLNM: occult cervical lymph node metastasis.

Categorical Variables

	n	n%		n	n%
Gender			**Growth pattern**		
Male	102	51.3%	*Endophytic*	112	72.3%
Female	97	48.7%	*Exophytic*	43	27.7%
NA	0		*NA*	44	
Smoking			**cT**		
Yes	104	53.6%	*1*	64	32.2%
No	90	46.4%	*2*	65	32.7%
NA	5		*3*	9	4.5%
			4	61	30.7%
Alcohol			*NA*	0	
Yes	86	44.6%			
No	107	55.4%	**pT**		
NA	6		*1*	79	39.9%
			2	46	23.2%
ACE27 Comorbidity			*3*	11	5.6%
None	93	47.4%	*4*	62	31.3%
Mild	56	28.6%	*NA*	1	
Moderate	34	17.3%			
Severe	13	6.6%	**cN**		
NA	3		*0*	136	68.3%
			1	60	30.2%
Institution			*2*	3	1.5%
JRH	177	88.9%	*3*	0	0.0%
UCLH	22	11.1%			
NA	0		**pN**		
			0	129	65.2%
Subsite			*1*	20	10.1%
Oral Tongue	97	48.7%	*2*	49	24.7%
Floor	31	15.6%	*3*	0	0.0%
Lower Alveolus	26	13.1%	*NA*	1	
Upper Alveolus	14	7.0%			
Hard Palate	2	1.0%	**cStage**		
Retromolar Trigone	18	9.0%	*I*	57	28.6%
Buccal	11	5.5%	*II*	48	24.1%
			III	33	16.6%
Location			*IV*	61	30.7%
Left	95	47.7%	*NA*	0	
Right	93	46.7%			
Midline	11	5.5%			
NA	0				

	n	n%
pStage		
I	65	32.8%
II	27	13.6%
III	20	10.1%
IV	86	43.4%
NA	1	
MRI CLNM		
Negative	82	44.8%
Equivocal	67	36.6%
Positive	34	18.6%
NA	16	
Pathology CLNM		
None	130	65.3%
Ipsilateral	63	31.7%
Bilateral	6	3.0%
NA	0	
Occult Metastasis		
No	110	81.5%
Yes	25	18.5%
NA	0	
Single/Multiple CLNM		
None	130	65.3%
Single	34	17.1%
Multiple	35	17.6%
NA		
CNLM Levels		
I	25	17.2%
II	37	25.5%
III	47	32.4%
IV	27	18.6%
V	9	6.2%
NA	0	
MRI ECS		
No	5	2.6%
Yes	194	97.4%
NA	161	

	n	n%
Pathology ECS		
No	151	75.9%
Yes	48	24.1%
NA	0	
Pathology ECS Levels		
I	24	28.9%
II	29	34.9%
III	21	25.3%
IV	7	8.4%
V	2	2.4%
NA	0	
Imaging Bone		
No	144	72.7%
Yes	54	27.3%
NA	1	
Pathology Bone		
No	152	77.2%
Yes	45	22.8%
NA	2	
Midline		
No	159	79.9%
Abuts	8	4.0%
Involved	32	16.1%
NA	0	
Grade		
Well	38	20.0%
Moderately	132	69.5%
Poorly	20	10.5%
NA	9	
Invasive Front		
Non-Cohesive	42	23.0%
Cohesive	141	77.0%
NA	16	
LVI		
No	156	79.2%
Yes	41	20.8%
NA	2	

	n	n%
PNI		
No	127	64.5%
Yes	70	35.5%
NA	2	
Surgical Margins		
Clear	90	45.9%
Close	84	42.9%
Involved	22	11.2%
NA	3	
Neck Dissection		
None	27	13.6%
SND(I-III)	16	8.1%
SND(I-IV)	45	22.7%
SND(NOS)	2	1.0%
MRND	81	40.9%
BND	27	13.6%
NA	1	
Bone Resection		
None	124	62.3%
Marginal	32	16.1%
Segmental	33	16.6%
Maxillectomy	10	5.0%
NA	0	
Reconstruction		
None	18	9.1%
Mucosal Graft	19	9.6%
Pedicled Flap	7	3.5%
Obturator	8	4.0%
Free Flap	146	73.7%
NA	1	
Adjuvant Therapy		
None	131	65.8%
Radiotherapy	58	29.1%
Chemotherapy	1	0.5%
Chemoradiotherapy	9	4.5%
NA	0	

	n	n%
Recurrence		
None	150	78.9%
Local	14	7.4%
Regional	17	8.9%
Distant	1	0.5%
Local and Regional	3	1.6%
Local and Distant	1	0.5%
Regional and Distant	4	2.1%
NA	0	
Second Primary		
No	186	97.9%
Oral tongue	1	0.5%
Buccal	1	0.5%
Tonsil	1	0.5%
UADC	1	0.5%
NA	0	
Deaths		
No	140	70.4%
Peri-operative	8	4.0%
SCC related	38	19.1%
SCC Unrelated	13	6.5%
NA	0	
MRI Volume (median)		
<= 3.0 cm^3	65	32.7%
> 3.0 cm^3	134	67.3%
NA	0	
MRI Volume (quartiles)		
<= 0.5 cm^3	37	18.6%
> 0.5, <= 8.0	70	35.2%
> 8.0 cm^3	92	46.2%
NA	0	
MRI Thickness (median)		
<= 11.0 mm	65	32.7%
> 11.0 mm	134	67.3%
NA	0	

	n	n%		n	n%
MRI Thickness (quartiles)			**Pathology Diameter (quartiles)**		
<= 4.0 mm	37	18.6%	*<= 10.0 mm*	33	16.6%
> 4.0, <= 16.0	81	40.7%	*>10.0, <=25.0*	84	42.2%
> 16.0 mm	81	40.7%	*> 25.0 mm*	72	36.2%
NA	0		*NA*	10	
MRI Diameter (median)			**Pathology Depth (median)**		
<= 18.5 mm	62	31.2%	*<= 6.0 mm*	76	38.2%
> 18.5 mm	127	63.8%	*> 6.0 mm*	115	57.8%
NA	0		*NA*	8	
MRI Diameter (quartiles)			**Pathology Depth (quartiles)**		
<= 0.0 mm	37	18.6%	*<= 3.0 mm*	33	16.6%
> 0.0, <= 28.5	93	46.7%	*> 3.0, <=10.0*	82	41.2%
> 28.5 mm	69	34.7%	*> 10.0 mm*	76	38.2%
NA	0		*NA*	8	
Pathology Diameter (median)					
<= 15.5 mm	62	31.2%			
> 15.5 mm	137	68.6%			
NA					

Continuous Variables

	Mean	95% CI	Median	IQR
Age	61.8	(59.7 - 63.9)	61.6	(52.5 - 73.6)
Pathological Diameter	25.1	(22.9 - 27.3)	22.0	(14.0 - 35.0)
Pathological Depth	10.6	(9.4 - 11.7)	10.0	(5.0 - 15.0)
Pathological Diameter ($log_{10}X$)	1.32	(1.28 - 1.37)	1.36	(1.18 - 1.56)
Pathological Depth ($log_{10}X$)	0.96	(0.92 - 1.01)	1.04	(0.78 - 1.20)
MRI Volume	7.5	(5.7 - 9.2)	3.0	(0.5 - 8.0)
MRI Volume ($log_{10}X$)	0.62	(0.55 - 0.69)	0.60	(0.17 - 0.95)
MRI Thickness	12.3	(10.9 - 13.7)	11.2	(4.4 - 16.4)
MRI Thickness ($log_{10}X$)	0.93	(0.86 - 1.00)	1.09	(0.73 - 1.24)
MRI Diameter	27.7	(24.9 - 30.6)	25.4	(15.0 - 37.5)
MRI Diameter ($log_{10}X$)	1.22	(1.13 - 1.31)	1.42	(1.20 - 1.59)

Appendix IV-B: Occult Metastasis

The results of occult cervical lymph node metastasis analysis. Two separate tables are given for categorical and continuous variables. Categorical variables with more than two categories were dummy coded with the first listed category as the reference. For each category listed are the number of positive cases (Positive), the number of negative cases (Negative), the percentage of positive cases (Pos%), hazard ratio (HR) with 95% confidence interval and the p-value (p). For continuous variables listed are the mean with 95% confidence interval, median with interquartile range (IQR), hazard ratio with 95% confidence interval and the p-value. Significant p-values (≤ 0.05) are highlighted in green while p-values showing trend towards significance ($0.05 < p \leq 0.10$) are highlighted in yellow. In one cases there were no events in a category and, thus, no statistics could be computed.

HR: hazard ratio, p: p-value; ACE: adult comorbidity scale; OC: oral cavity; MRI: magnetic resonance imaging; CLN: cervical lymph node; ECS: extracapsular spread; PVI: perivascular invasion; PNI: perineural invasion; JRH: John Radcliffe Hospital; UCLH: University College London Hospital.

Categorical Variables

	Negative	Positive	Pos%	HR	95% CI	p
Sex						
Male	57	11	16.2%			
Female	54	14	20.6%	1.34	(0.56 - 3.22)	0.508
NA	0					
Location						
JRH	98	23	19.0%			
UCLH	13	2	13.3%	0.66	(0.14 - 3.11)	0.595
NA	0					
Smoking						
No	52	8	13.3%			
Yes	57	16	21.9%	1.82	(0.72 - 4.62)	0.204
NA	3					
Alcohol						
No	57	14	19.7%			
Yes	51	10	16.4%	0.80	(0.33 - 1.95)	0.622
NA	4					
ACE27						
None	45	13	22.4%			
Mild	38	5	11.6%	0.46	(0.15 - 1.39)	0.168
Moderate	23	3	11.5%	0.45	(0.12 - 1.75)	0.249
Severe	5	4	44.4%	2.77	(0.65 - 11.83)	0.169
NA	0					
Site						
Oral Tongue	60	12	16.7%			
Lower OC	33	11	25.0%	1.67	(0.66 - 4.19)	0.277
Upper OC	18	2	10.0%	0.56	(0.11 - 2.72)	0.467
NA	0					
Subsite						
Left	51	12	19.0%			
Midline	7	0	0.0%	NA	NA	NA
Right	53	13	19.7%	1.04	(0.44 - 2.5)	0.926
NA	0					
cT						
1	50	7	12.3%			
2	40	8	16.7%	1.43	(0.48 - 4.28)	0.523
3	3	1	25.0%	2.38	(0.22 - 26.18)	0.478
4	18	9	33.3%	3.57	(1.16 - 11)	0.027
NA	0					
Pattern						
Endophytic	52	16	23.5%			
Exophytic	30	4	11.8%	0.43	(0.13 - 1.42)	0.166
NA	34					

	Negative	Positive	Pos%	HR	95% CI	p
Midline						
Uninvolved	97	20	17.1%			
Abuts	10	3	23.1%	1.45	(0.37 - 5.77)	0.594
Crosses	4	2	33.3%	2.42	(0.42 - 14.16)	0.325
NA	0					
Imaging Bone						
No	93	16	14.7%			
Yes	18	9	33.3%	2.91	(1.11 - 7.59)	0.029
NA	0					
pT						
1	63	7	10.0%			
2	24	7	22.6%	2.62	(0.83 - 8.28)	0.100
3	4	2	33.3%	4.50	(0.69 - 29.15)	0.115
4	19	9	32.1%	4.26	(1.4 - 12.97)	0.011
NA	1					
Grade						
Well	25	2	7.4%			
Moderate	72	19	20.9%	3.30	(0.72 - 15.18)	0.125
Poor	6	4	40.0%	8.33	(1.23 - 56.67)	0.030
NA	8					
Front						
Cohesive	27	4	12.9%			
Non-Cohesive	71	20	22.0%	1.90	(0.6 - 6.07)	0.278
NA	14					
PVI						
No	100	16	13.8%			
Yes	10	9	47.4%	5.62	(1.98 - 15.97)	0.001
NA	1					
PNI						
No	83	14	14.4%			
Yes	27	11	28.9%	2.42	(0.98 - 5.95)	0.055
NA	1					
pBone						
No	93	19	17.0%			
Yes	17	6	26.1%	1.73	(0.6 - 4.95)	0.309
NA	1					
MRI Volume (median)						
<= 3.0 cm^3	78	8	9.3%			
> 3.0 cm^3	33	17	34.0%	5.02	(1.97 - 12.78)	<0.001
NA	0					

	Negative	Positive	Pos%	HR	95% CI	p
MRI Volume (quartiles)						
<= 0.5 cm^3	45	2	4.3%			
> 0.5, <= 8.0	52	16	23.5%	6.92	(1.51 - 31.78)	0.013
> 8.0 cm^3	14	7	33.3%	11.25	(2.09 - 60.49)	0.005
NA						
MRI Thickness (median)						
<= 11.0 mm	70	7	9.1%			
> 11.0 mm	41	18	30.5%	4.39	(1.69 - 11.40)	0.002
NA	0					
MRI Thickness (quartiles)						
<= 4.0 mm	5	2	5.0%			
> 4.0, <= 16.0	57	13	18.6%	4.33	(0.93 - 20.30)	0.063
> 16.0 mm	16	10	38.5%	11.88	(2.33 - 60.41)	0.003
NA	0					
MRI Diameter (median)						
<= 18.5 mm	55	4	6.8%			
> 18.5 mm	56	21	27.2%	5.16	(1.66 - 16.00)	0.005
NA	0					
MRI Diameter (quartiles)						
<= 0.0 mm	31	2	6.1%			
> 0.0, <= 28.5	52	8	13.3%	2.39	(0.48 - 11.95)	0.291
> 28.5 mm	28	15	34.9%	8.30	(1.74 - 39.6)	0.008
NA	0					
Pathology Diameter (median)						
<= 15.5 mm	52	5	8.8%			
> 15.5 mm	59	20	25.3%	3.53	(1.24 - 10.06)	0.019
NA	0					
Pathology Diameter (quartiles)						
<= 10.0 mm	28	2	6.7%			
>10.0, <=25.0	54	12	18.2%	3.11	(0.65 - 14.88)	0.155
> 25.0 mm	19	11	27.5%	5.31	(1.08 - 26.14)	0.040
NA	0					
Pathology Depth (median)						
<= 6.0 mm	60	9	13.0%			
> 6.0 mm	51	16	23.9%	2.09	(0.85 - 5.13)	0.107
NA	0					
Pathology Depth (quartiles)						
<= 3.0 mm	29	2	6.5%			
> 3.0, <=10.0	51	12	19.0%	3.41	(0.71 - 16.31)	0.124
> 10.0 mm	31	11	26.2%	5.15	(1.05 - 25.22)	0.043
NA	0					

Continuous Variables

	Mean	95% CI	Median	IQR	HR	95% CI	p
Age							
No	62.1	(59.2 - 65.1)	61.4	(52.5 - 73.7)			
Yes	61.6	(54.7 - 68.5)	62.8	(50.8 - 75.0)	1.00	(0.97 - 1.03)	0.815
Pathological Diameter							
No	18.9	(16.5 - 21.3)	15.5	(10.0 - 25.0)			
Yes	28.3	(22.3 - 34.3)	25.0	(18.0 - 38.5)	1.04	(1.01 - 1.07)	0.007
Pathological Depth							
No	7.8	(6.6 - 8.9)	6.0	(3.0 - 11.0)			
Yes	11.6	(8.2 - 15.0)	10.0	(5.3 - 16.0)	13.85	(2.19 - 87.63)	0.005
Pathological Diameter ($log_{10}X$)							
No	1.20	(1.14 - 1.27)	1.22	(1.04 - 1.42)			
Yes	1.41	(1.31 - 1.51)	1.42	(1.28 - 1.60)	1.08	(1.02 - 1.15)	0.010
Pathological Depth ($log_{10}X$)							
No	0.84	(0.78 - 0.90)	0.85	(0.60 - 1.08)			
Yes	1.02	(0.91 - 1.13)	1.04	(0.80 - 1.23)	9.43	(1.90 - 46.89)	0.006
MRI Volume							
No	3.8	(1.9 - 5.6)	1.0	(0.0 - 3.6)			
Yes	8.4	(4.4 - 12.4)	4.1	(2.5 - 11.0)	1.04	(1.00 - 1.08)	0.060
MRI Volume ($log_{10}X$)							
No	0.41	(0.33 - 0.50)	0.31	(0.00 - 0.66)			
Yes	0.78	(0.60 - 0.95)	0.71	(0.54 - 1.08)	6.18	(2.25 - 16.99)	<0.001
MRI Thickness							
No	9.1	(7.4 - 10.8)	7.6	(0.0 - 13.9)			
Yes	15.0	(11.6 - 18.5)	14.0	(10.8 - 18.2)	1.07	(1.02 - 1.13)	0.003
MRI Thickness ($log_{10}X$)							
No	0.78	(0.68 - 0.88)	0.93	(0.00 - 1.17)			
Yes	1.11	(0.95 - 1.26)	1.18	(1.07 - 1.28)	6.30	(1.79 - 22.19)	0.004
MRI Diameter							
No	20.0	(16.8 - 23.2)	19.3	(0.0 - 30.0)			
Yes	34.2	(26.2 - 42.1)	29.6	(23.8 - 46.1)	1.04	(1.02 - 1.07)	<0.001
MRI Diameter ($log_{10}X$)							
No	1.05	(0.92 - 1.17)	1.31	(0.00 - 1.49)			
Yes	1.42	(1.23 - 1.61)	1.49	(1.39 - 1.67)	4.80	(1.44 - 15.99)	0.011

Appendix IV-C: All-Cause Survival

The results of all-cause survival analysis. Two separate tables are given for categorical and continuous variables. Categorical variables with more than two categories were dummy coded with the first listed category as the reference. For each category listed are the number of events (E), the number of censored cases (C), the percentage of cases with events (E%), the mean with 95% confidence interval (CI), hazard ratio (HR) with 85% confidence interval and the p-value (p). For continuous variables listed are the mean, median and hazard ratio, all with 95% confidence intervals, and the p-value. Significant p-values (≤ 0.05) are highlighted in green while p-values showing trend towards significance ($0.05 < p \leq 0.10$) are highlighted in yellow. In three cases, there were no events in the reference category of dummy-coded variables and, thus, no statistics could be computed.

E: events; C: censored; CI: confidence interval; HR: hazard ratio, p: p-value; ACE: adult comorbidity scale; OC: oral cavity; MRI: magnetic resonance imaging; CLN: cervical lymph node; ECS: extracapsular spread; PVI: perivascular invasion; PNI: perineural invasion; JRH: John Radcliffe Hospital; UCLH: University College London Hospital.

Categorical Variables

	E	C	E%	Mean	95% CI	HR	95% CI	p
Gender								
Male	13	83	13.5%	668	(636 - 700)			
Female	20	74	21.3%	629	(588 - 671)	1.68	(0.84 - 3.38)	0.144
Smoking								
No	14	71	16.5%	654	(616 - 692)			
Yes	19	82	18.8%	641	(604 - 679)	1.20	(0.6 - 2.39)	0.609
Alcohol								
No	20	82	19.6%	638	(601 - 676)			
Yes	13	70	15.7%	658	(620 - 696)	0.82	(0.41 - 1.64)	0.568
ACE27								
None	13	76	14.6%	669	(637 - 701)			
Mild	9	49	15.5%	647	(595 - 698)	1.24	(0.53 - 2.90)	0.622
Moderate	8	24	25.0%	620	(547 - 692)	1.96	(0.81 - 4.73)	0.134
Severe	3	8	27.3%	584	(448 - 720)	2.47	(0.70 - 8.71)	0.158
Institution								
JRH	29	141	17.1%	651	(624 - 679)			
UCLH	4	16	20.0%	634	(549 - 719)	1.22	(0.43 - 3.46)	0.713
Subsite (Combined)								
Oral Tongue	14	79	15.1%	660	(625 - 696)			
Lower OC	14	57	19.7%	631	(584 - 678)	1.42	(0.68 - 2.99)	0.350
Upper OC	5	21	19.2%	656	(593 - 720)	1.22	(0.44 - 3.38)	0.705
Location								
Left	21	70	23.1%	630	(590 - 670)			
Midline	1	9	10.0%	674	(570 - 778)	0.47	(0.06 - 3.51)	0.464
Right	11	78	12.4%	668	(632 - 703)	0.55	(0.27 - 1.15)	0.113
cT								
1	6	57	9.5%	684	(649 - 719)			
2	10	53	15.9%	649	(601 - 698)	1.69	(0.61 - 4.66)	0.309
3	2	7	22.2%	634	(514 - 754)	2.10	(0.42 - 10.42)	0.363
4	15	40	27.3%	615	(562 - 667)	2.94	(1.14 - 7.57)	0.026
cN								
0	12	120	9.1%	691	(669 - 713)			
1	19	36	34.5%	565	(502 - 629)	4.40	(2.14 - 9.07)	<0.001
2	2	1	66.7%	263	(209 - 316)	25.61	(5.37 - 122.25)	<0.001
cStage								
1	3	54	5.3%	706	(680 - 733)			
2	3	43	6.5%	695	(656 - 735)	1.19	(0.24 - 5.88)	0.835
3	12	20	37.5%	548	(463 - 634)	8.10	(2.29 - 28.72)	0.001
4	15	40	27.3%	615	(562 - 667)	5.39	(1.56 - 18.62)	0.008

	E	C	E%	Mean	95% CI	HR	95% CI	p
Growth Pattern								
Endophytic	21	84	20.0%	631	(592 - 670)			
Exophytic	8	34	19.0%	655	(602 - 708)	0.84	(0.37 - 1.90)	0.679
MRI CLNM								
No	18	127	12.4%	672	(645 - 698)			
Yes	10	22	31.3%	598	(523 - 673)	2.65	(1.22 - 5.74)	0.014
MRI ECS								
No	26	148	14.9%	662	(637 - 688)			
Yes	2	2	50.0%	491	(245 - 737)	3.77	(0.89 - 15.89)	0.071
Midline								
Uninvolved	23	131	14.9%	658	(630 - 686)			
Abuts	6	23	20.7%	643	(576 - 709)	1.39	(0.57 - 3.42)	0.468
Crosses	4	3	57.1%	492	(317 - 667)	4.41	(1.52 - 12.77)	0.006
Imaging Bone								
No	19	120	13.7%	664	(635 - 693)			
Yes	14	36	28.0%	608	(551 - 665)	2.13	(1.07 - 4.24)	0.032
pT								
1	5	72	6.5%	697	(669 - 726)			
2	8	38	17.4%	650	(595 - 705)	2.89	(0.95 - 8.84)	0.063
3	2	8	20.0%	632	(507 - 756)	3.13	(0.61 - 16.15)	0.172
4	18	38	32.1%	585	(527 - 642)	5.67	(2.10 - 15.29)	0.001
pN								
0	8	118	6.3%	707	(690 - 724)			
1	8	11	42.1%	551	(451 - 650)	7.94	(2.97 - 21.19)	<0.001
2	17	26	39.5%	520	(443 - 598)	8.78	(3.77 - 20.42)	<0.001
pStage								
1	1	64	1.5%	724	(710 - 737)			
2	3	24	11.1%	679	(619 - 739)	7.48	(0.78 - 71.94)	0.081
3	4	14	22.2%	645	(563 - 727)	15.13	(1.69 - 135.39)	0.015
4	25	53	32.1%	576	(525 - 627)	25.42	(3.44 - 187.71)	0.002
Grade								
Well	5	33	13.2%	683	(639 - 727)			
Moderate	19	106	15.2%	660	(630 - 690)	1.19	(0.44 - 3.18)	0.735
Poor	9	10	47.4%	484	(366 - 602)	4.90	(1.64 - 14.65)	0.004
Invasive Front								
Cohesive	3	37	7.5%	700	(664 - 735)			
Non-Cohesive	28	106	20.9%	631	(597 - 665)	3.21	(0.98 - 10.57)	0.055
Surgical Margin								
Clear	11	76	12.6%	673	(641 - 705)			
Close	16	63	20.3%	631	(586 - 677)	1.66	(0.77 - 3.59)	0.193
Involved	5	16	23.8%	629	(544 - 713)	1.94	(0.68 - 5.60)	0.218

	E	C	E%	Mean	95% CI	HR	95% CI	p
PVI								
No	20	130	13.3%	673	(647 - 698)			
Yes	13	25	34.2%	542	(462 - 623)	3.54	(1.75 - 7.15)	<0.001
PNI								
No	17	107	13.7%	668	(639 - 696)			
Yes	16	48	25.0%	608	(554 - 662)	2.09	(1.06 - 4.15)	0.034
Pathology Bone								
No	20	126	13.7%	663	(635 - 692)			
Yes	13	30	30.2%	598	(536 - 660)	2.42	(1.20 - 4.87)	0.013
Pathology ECS								
No	18	129	12.2%	679	(655 - 703)			
Yes	15	28	34.9%	551	(477 - 625)	3.47	(1.74 - 6.91)	<0.001
Positive CLNM								
No	8	118	6.3%	707	(690 - 724)			
Yes	25	37	40.3%	530	(469 - 592)	8.49	(3.82 - 18.88)	<0.001
Type of CLNM								
None	8	118	6.3%	707	(690 - 724)			
Occult	5	17	22.7%	619	(534 - 703)	4.34	(1.41 - 13.31)	0.010
Clinical	20	21	48.8%	483	(404 - 562)	11.01	(4.83 - 25.07)	<0.001
MRI Volume (median)								
<= 3.0 cm^3	9	88	9.3%	683	(653 - 713)			
> 3.0 cm^3	24	69	25.8%	615	(573 - 657)	2.92	(1.35 - 6.27)	0.006
MRI Volume (quartiles)								
<= 0.5 cm^3	2	49	3.9%	711	(683 - 738)			
> 0.5, <= 8.0	15	79	16.0%	648	(609 - 687)	4.45	(1.02 - 19.47)	0.047
> 8.0 cm^3	16	29	35.6%	581	(517 - 644)	10.74	(2.47 - 46.76)	0.002
MRI Thickness (median)								
<= 11.0 mm	10	84	10.6%	674	(641 - 707)			
> 11.0 mm	23	73	24.0%	625	(585 - 664)	2.34	(1.11 - 4.92)	0.025
MRI Thickness (quartiles)								
<= 4.0 mm	1	42	97.7%	718	(693 - 743)			
> 4.0, <= 16.0	14	75	84.3%	652	(614 - 690)	7.32	(0.96 - 55.67)	0.054
> 16.0 mm	18	40	69.0%	598	(543 - 653)	13.97	(1.86 - 104.67)	0.010
MRI Diameter (median)								
<= 18.5 mm	2	61	3.2%	714	(691 - 737)			
> 18.5 mm	31	96	24.4%	618	(582 - 654)	8.44	(2.02 - 35.30)	0.003
MRI Diameter (quartiles)								
<= 0.0 mm	0	36	NA	NA	NA			
> 0.0, <= 18.5	11	59	NA	NA	NA	NA	NA	NA
> 28.5 mm	22	62	NA	NA	NA	NA	NA	NA

	E	C	E%	Mean	95% CI	HR	95% CI	p
Pathology Diameter (median)								
<= *15.5 mm*	3	57	5.0%	708	(683 - 733)			
> *15.5 mm*	28	93	23.1%	621	(584 - 658)	5.26	(1.6 - 17.31)	0.006
Pathology Diameter (quartiles)								
<= *10.0 mm*	2	31	93.9%	701	(661 - 741)			
> *10.0,<= 25.0*	9	72	88.9%	678	(645 - 712)	1.86	(0.4 - 8.61)	0.428
> *25.0 mm*	20	48	70.6%	591	(537 - 645)	5.67	(1.32 - 24.27)	0.019
Pathology Depth (median)								
<= *6.0 mm*	4	70	5.4%	704	(678 - 729)			
> *6.0 mm*	29	81	26.4%	608	(568 - 649)	5.23	(1.84 - 14.89)	0.002
Pathology Depth (quartiles)								
<= *3.0 mm*	0	32	NA	NA	NA			
> *3.0, <= 10.0*	11	68	NA	NA	NA	NA	NA	NA
> *10.0 mm*	22	52	NA	NA	NA	NA	NA	NA

Continuous Variables

	Mean	95% CI	Median	IQR	HR	95% CI	p
Age							
Censored	60.2	(57.9 - 62.5)	60.0	(50.6 - 69.6)	1.02	(1 - 1.05)	0.063
Event	65.0	(59.6 - 70.4)	64.9	(56.3 - 77.7)			
Pathology Diameter							
Censored	23.2	(20.8 - 25.6)	20.0	(12.0 - 31.0)	1.03	(1.02 - 1.05)	<0.001
Event	35.3	(28.7 - 41.9)	35.0	(20.0 - 45.0)			
Pathology Depth							
Censored	9.3	(8.1 - 10.4)	7.0	(4.0 - 13.0)	1.06	(1.02 - 1.1)	0.002
Event	14.1	(11.3 - 17.0)	12.6	(8.0 - 18.0)			
Pathology Diameter ($log_{10}X$)							
Censored	1.29	(1.23 - 1.35)	1.30	(1.10 - 1.54)	5.74	(1.95 - 16.87)	0.001
Event	1.51	(1.42 - 1.61)	1.56	(1.30 - 1.69)			
Pathology Depth ($log_{10}X$)							
Censored	0.86	(0.79 - 0.94)	0.90	(0.60 - 1.18)	3.24	(1.55 - 6.78)	0.002
Event	1.15	(1.03 - 1.26)	1.13	0.93 - 1.30)			
MRI Volume							
Censored	5.4	(3.8 - 7.0)	2.0	(0.3 - 5.9)	1.03	(1.01 - 1.05)	<0.001
Event	13.7	(8.7 - 18.7)	7.7	(2.6 - 22.0)			
MRI Volume ($log_{10}X$)							
Censored	0.53	(0.46 - 0.60)	0.48	(0.12 - 0.83)	4.45	(2.26 - 8.73)	<0.001
Event	0.96	(0.79 - 1.12)	0.94	(0.56 - 1.35)			
MRI Thickness							
Censored	10.5	(9.1 - 11.9)	10.5	(3.2 - 15.5)	1.06	(1.03 - 1.09)	<0.001
Event	17.7	(14.2 - 21.2)	16.3	(9.5 - 27.1)			
MRI Thickness ($log_{10}X$)							
Censored	0.90	(0.80 - 0.90)	1.06	(0.60 - 1.20)	7.46	(2.26 - 24.59)	<0.001
Event	1.20	(1.10 - 1.30)	1.24	(1.00 - 1.40)			
MRI Diameter							
Censored	24.4	(21.3 - 27.5)	22.9	(12.5 - 35.0)	1.03	(1.01 - 1.04)	<0.001
Event	39.5	(34.1 - 44.9)	37.5	(25.9 - 52.2)			
MRI Diameter ($log_{10}X$)							
Censored	1.10	(1.00 - 1.20)	1.38	(1.10 - 1.60)	12.54	(2.85 - 55.13)	<0.001
Event	1.60	(1.50 - 1.60)	1.59	(1.40 - 1.70)			

Appendix IV-D: Disease-Related Survival

The results of all-cause survival analysis. Two separate tables are given for categorical and continuous variables. Categorical variables with more than two categories were dummy coded with the first listed category as the reference. For each category listed are the number of events (E), the number of censored cases (C), the percentage of cases with events (E%), the mean with 95% confidence interval (CI), hazard ratio (HR) with 85% confidence interval and the p-value (p). For continuous variables listed are the mean, median and hazard ratio, all with 95% confidence intervals, and the p-value. Significant p-values (≤ 0.05) are highlighted in green while p-values showing trend towards significance ($0.05 < p \leq 0.10$) are highlighted in yellow. In three cases, there were no events in the reference category of dummy-coded variables and, thus, no statistics could be computed.

E: events; C: censored; CI: confidence interval; HR: hazard ratio, p: p-value; ACE: adult comorbidity scale; OC: oral cavity; MRI: magnetic resonance imaging; CLN: cervical lymph node; ECS: extracapsular spread; PVI: perivascular invasion; PNI: perineural invasion; JRH: John Radcliffe Hospital; UCLH: University College London Hospital.

Continuous Variables

	E	C	E%	Mean	95% CI	HR	95% CI	p
Gender								
Male	11	85	11.5%	676	(645 - 706)			
Female	18	76	19.1%	638	(598 - 678)	1.80	(0.85 - 3.82)	0.124
Smoking								
No	13	72	15.3%	657	(618 - 695)			
Yes	16	85	15.8%	654	(618 - 689)	1.09	(0.52 - 2.26)	0.820
Alcohol								
No	19	83	18.6%	641	(603 - 678)			
Yes	10	73	12.0%	674	(640 - 708)	0.66	(0.31 - 1.43)	0.293
ACE27								
None	13	76	14.6%	669	(637 - 701)			
Mild	7	51	12.1%	661	(613 - 710)	0.97	(0.39 - 2.43)	0.946
Moderate	6	26	18.8%	642	(574 - 710)	1.48	(0.56 - 3.90)	0.426
Severe	3	8	27.3%	584	(448 - 720)	2.44	(0.69 - 8.57)	0.165
Institution								
JRH	25	145	14.7%	660	(634 - 686)			
UCLH	4	16	20.0%	634	(549 - 719)	1.40	(0.49 - 4.02)	0.532
Subsite (Combined)								
Oral Tongue	12	81	12.9%	669	(636 - 703)			
Lower OC	12	59	16.9%	641	(595 - 687)	1.42	(0.64 - 3.16)	0.390
Upper OC	5	21	19.2%	656	(593 - 720)	1.43	(0.50 - 4.06)	0.502
Location								
Left	19	72	20.9%	639	(600 - 678)			
Midline	1	9	10.0%	674	(570 - 778)	0.53	(0.07 - 3.96)	0.536
Right	9	80	10.1%	676	(642 - 710)	0.50	(0.23 - 1.11)	0.087
cT								
1	5	58	7.9%	692	(660 - 724)			
2	9	54	14.3%	658	(613 - 704)	1.85	(0.62 - 5.53)	0.269
3	2	7	22.2%	634	(514 - 754)	2.54	(0.49 - 13.10)	0.265
4	13	42	23.6%	623	(570 - 676)	3.06	(1.09 - 8.60)	0.033
cN								
0	8	124	6.1%	703	(685 - 722)			
1	19	36	34.5%	565	(502 - 629)	6.63	(2.90 - 15.14)	<0.001
2	2	1	66.7%	263	(209 - 316)	39.45	(7.77 - 200.33)	<0.001

	E	C	E%	Mean	95% CI	HR	95% CI	p
cStage								
1	2	55	3.5%	715	(695 - 736)			
2	2	44	4.3%	709	(680 - 739)	1.21	(0.17 - 8.56)	0.852
3	12	20	37.5%	548	(463 - 634)	12.22	(2.73 - 54.59)	0.001
4	13	42	23.6%	623	(570 - 676)	7.03	(1.59 - 31.17)	0.010
Growth Pattern								
Endophytic	20	85	19.0%	635	(597 - 674)			
Exophytic	5	37	11.9%	680	(634 - 725)	0.55	(0.21 - 1.48)	0.237
MRI CLNM								
No	14	131	9.7%	682	(658 - 707)			
Yes	10	22	31.3%	598	(523 - 673)	3.43	(1.52 - 7.73)	0.003
MRI ECS								
No	22	152	12.6%	671	(647 - 695)			
Yes	2	2	50.0%	491	(245 - 737)	4.47	(1.05 - 19.02)	0.043
Midline								
Uninvolved	20	134	13.0%	667	(640 - 693)			
Abuts	6	23	20.7%	643	(576 - 709)	1.59	(0.64 - 3.96)	0.319
Crosses	3	4	42.9%	521	(335 - 708)	3.81	(1.13 - 12.84)	0.031
Imaging Bone								
No	16	123	11.5%	673	(646 - 700)			
Yes	13	37	26.0%	613	(556 - 670)	2.33	(1.12 - 4.84)	0.024
pT								
1	5	72	6.5%	697	(669 - 726)			
2	5	41	10.9%	679	(633 - 725)	1.80	(0.52 - 6.23)	0.351
3	2	8	20.0%	632	(507 - 756)	3.09	(0.60 - 15.94)	0.177
4	17	39	30.4%	589	(531 - 647)	5.31	(1.96 - 14.39)	0.001
pN								
0	4	122	3.2%	720	(709 - 731)			
1	8	11	42.1%	551	(451 - 650)	15.77	(4.74 - 52.45)	<0.001
2	17	26	39.5%	520	(443 - 598)	17.22	(5.78 - 51.30)	<0.001
pStage								
1	1	64	1.5%	724	(710 - 737)			
2	0	27	0.0%	NA	NA	NA	NA	NA
3	4	14	22.2%	645	(563 - 727)	14.91	(1.67 - 133.39)	0.016
4	24	54	30.8%	579	(528 - 631)	24.17	(3.27 - 178.75)	0.002
Grade								
Well	4	34	10.5%	689	(646 - 732)			
Moderate	16	109	12.8%	670	(642 - 698)	1.25	(0.42 - 3.73)	0.695
Poor	9	10	47.4%	484	(366 - 602)	6.06	(1.86 - 19.71)	0.003
Invasive Front								
Cohesive	1	39	2.5%	718	(693 - 742)			
Non-Cohesive	26	108	19.4%	637	(604 - 670)	8.92	(1.21 - 65.74)	0.032

	E	C	E%	Mean	95% CI	HR	95% CI	p
Surgical Margin								
Clear	9	78	10.3%	681	(651 - 712)			
Close	14	65	17.7%	641	(598 - 685)	1.79	(0.78 - 4.14)	0.172
Involved	5	16	23.8%	629	(544 - 713)	2.37	(0.79 - 7.06)	0.123
PVI								
No	16	134	10.7%	683	(660 - 706)			
Yes	13	25	34.2%	542	(462 - 623)	4.35	(2.08 - 9.09)	<0.001
PNI								
No	15	109	12.1%	673	(646 - 701)			
Yes	14	50	21.9%	621	(569 - 673)	2.06	(0.99 - 4.27)	0.052
Pathology Bone								
No	18	128	12.3%	669	(642 - 696)			
Yes	11	32	25.6%	614	(552 - 675)	2.27	(1.07 - 4.81)	0.032
Pathology ECS								
No	14	133	9.5%	690	(668 - 711)			
Yes	15	28	34.9%	551	(477 - 625)	4.39	(2.11 - 9.11)	<0.001
Positive CLNM								
No	4	122	3.2%	720	(709 - 731)			
Yes	25	37	40.3%	530	(469 - 592)	16.72	(5.81 - 48.16)	<0.001
Type of CLNM								
None	4	122	3.2%	720	(709 - 731)			
Occult	5	17	22.7%	619	(534 - 703)	8.44	(2.26 - 31.50)	0.002
Clinical	20	21	48.8%	483	(404 - 562)	21.92	(7.48 - 64.28)	<0.001
MRI Volume (median)								
<= 3.0 cm^3	7	90	7.2%	695	(669 - 720)			
> 3.0 cm^3	22	71	23.7%	620	(578 - 662)	3.41	(1.46 - 7.98)	0.005
MRI Volume (quartiles)								
<= 0.5 cm^3	1	50	98.0%	721	(702 - 740)			
> 0.5, <= 8.0	14	80	85.1%	0	(0 - 0)	8.22	(1.08 - 62.53)	0.042
> 8.0 cm^3	14	31	68.9%	0	(0 - 0)	18.55	(2.44 - 141.17)	0.005
MRI Thickness (median)								
<= 11.0 mm	8	86	8.5%	686	(656 - 716)			
> 11.0 mm	21	75	21.9%	630	(590 - 669)	2.65	(1.18 - 6)	0.019
MRI Thickness (quartiles)								
<= 4.0 mm	0	43	NA	NA	NA			
> 4.0, <= 16.0	12	77	NA	NA	NA	NA	NA	NA
> 16.0 mm	17	41	NA	NA	NA	NA	NA	NA
MRI Diameter (median)								
<= 18.5 mm	2	61	3.2%	714	(691 - 737)			
> 18.5 mm	27	100	21.2%	630	(595 - 664)	7.28	(1.73 - 30.64)	0.007

	E	C	E%	Mean	95% CI	HR	95% CI	p
MRI Diameter (quartiles)								
<= 0.0 mm	0	36	NA	NA	NA			
> 0.0, <= 28.5	9	61	NA	NA	NA	NA	NA	NA
> 28.5 mm	20	64	NA	NA	NA	NA	NA	NA
Pathology Diameter (median)								
<= 15.5 mm	3	57	5.0%	708	(683 - 733)			
> 15.5 mm	24	97	19.8%	633	(597 - 669)	4.48	(1.35 - 14.88)	0.014
Pathology Diameter (quartiles)								
<= 10.0 mm	2	31	93.9%	701	(661 - 741)			
> 10.0,<= 25.0	7	74	91.4%	687	(656 - 718)	1.45	(0.3 - 6.96)	0.646
> 25.0 mm	18	50	73.5%	603	(550 - 656)	5.04	(1.17 - 21.73)	0.030
Pathology Depth (median)								
<= 6.0 mm	3	71	4.1%	711	(689 - 733)			
> 6.0 mm	26	84	23.6%	618	(578 - 657)	6.22	(1.88 - 20.56)	0.003
Pathology Depth (quartiles)								
<= 3.0 mm	0	32	NA	NA	NA			
> 3.0, <= 10.0	9	70	NA	NA	NA	NA	NA	NA
> 10.0 mm	20	54	NA	NA	NA	NA	NA	NA

Continuous Variables

	Mean	95% CI	Median	IQR	HR	95% CI	p
Age							
Censored	60.7	(58.4 - 63.0)	50.8	(61.0 - 70.7)	1.01	(0.99 - 1.04)	0.306
Event	63.0	(57.2 - 68.7)	53.4	(62.5 - 75.4)			
Pathology Diameter							
Censored	23.1	(20.8 - 25.4)	12.0	(20.0 - 32.0)	1.04	(1.02 - 1.05)	<0.001
Event	33.9	(27.8 - 40.0)	20.0	(36.0 - 45.0)			
Pathology Depth							
Censored	9.5	(8.4 - 10.7)	4.0	(8.0 - 13.5)	1.06	(1.02 - 1.11)	0.002
Event	14.9	(11.5 - 18.2)	9.5	(12.6 - 19.0)			
Pathology Diameter ($log_{10}X$)							
Censored	1.30	(1.2 - 1.31)	1.08	(1.30 - 1.51)	6.03	(1.9 - 19.13)	0.002
Event	1.50	(1.37 - 1.58)	1.30	(1.56 - 1.65)			
Pathology Depth ($log_{10}X$)							
Censored	0.80	(0.78 - 0.9)	0.60	(0.90 - 1.13)	3.50	(1.59 - 7.7)	0.002
Event	1.10	(1.02 - 1.21)	0.98	(1.10 - 1.28)			
MRI Volume							
Censored	5.4	(3.9 - 7.0)	0.3	(2.1 - 6.2)	1.03	(1.02 - 1.05)	<0.001
Event	14.6	(9.0 - 20.2)	3.3	(7.8 - 23.1)			
MRI Volume ($log_{10}X$)							
Censored	0.50	(0.46 - 0.61)	0.11	(0.49 - 0.86)	4.95	(2.4 - 10.19)	<0.001
Event	1.00	(0.81 - 1.16)	0.63	(0.94 - 1.38)			
MRI Thickness							
Censored	10.6	(9.2 - 12.0)	3.7	(10.5 - 15.5)	1.06	(1.03 - 1.1)	<0.001
Event	18.4	(14.6 - 22.1)	9.5	(16.5 - 27.5)			
MRI Thickness ($log_{10}X$)							
Censored	0.90	(0.78 - 0.94)	0.67	(1.06 - 1.22)	9.53	(2.46 - 36.88)	0.001
Event	1.20	(1.14 - 1.32)	1.02	(1.24 - 1.45)			
MRI Diameter							
Censored	24.6	(21.6 - 27.7)	12.7	(22.9 - 35.0)	1.03	(1.01 - 1.04)	<0.001
Event	40.4	(34.6 - 46.2)	27.4	(40.0 - 52.2)			
MRI Diameter ($log_{10}X$)							
Censored	1.10	(1.04 - 1.24)	1.14	(1.38 - 1.56)	14.84	(2.93 - 75.13)	0.001
Event	1.60	(1.52 - 1.65)	1.45	(1.61 - 1.73)			

Appendix IV-E: Disease-Free Survival

The results of all-cause survival analysis. Two separate tables are given for categorical and continuous variables. Categorical variables with more than two categories were dummy coded with the first listed category as the reference. For each category listed are the number of events (E), the number of censored cases (C), the percentage of cases with events (E%), the mean with 95% confidence interval (CI), hazard ratio (HR) with 85% confidence interval and the p-value (p). For continuous variables listed are the mean, median and hazard ratio, all with 95% confidence intervals, and the p-value. Significant p-values (≤ 0.05) are highlighted in green while p-values showing trend towards significance ($0.05 < p \leq 0.10$) are highlighted in yellow.

E: events; C: censored; CI: confidence interval; HR: hazard ratio, p: p-value; ACE: adult comorbidity scale; OC: oral cavity; MRI: magnetic resonance imaging; CLNM: cervical lymph node; ECS: extracapsular spread; PVI: perivascular invasion; PNI: perineural invasion; JRH: John Radcliffe Hospital; UCLH: University College London Hospital.

Categorical Variables

	E	C	E%	Mean	95% CI	HR	95% CI	p
Gender								
Male	18	78	18.8%	636	(596 - 676)			
Female	26	68	27.7%	590	(542 - 638)	1.58	(0.87 - 2.89)	0.134
Smoking								
No	16	69	18.8%	630	(584 - 675)			
Yes	27	74	26.7%	599	(555 - 643)	1.54	(0.83 - 2.85)	0.173
Alcohol								
No	22	80	21.6%	614	(571 - 658)			
Yes	21	62	25.3%	610	(563 - 657)	1.23	(0.67 - 2.23)	0.505
ACE27								
None	18	71	20.2%	636	(595 - 676)			
Mild	13	45	22.4%	603	(541 - 665)	1.31	(0.64 - 2.68)	0.457
Moderate	9	23	28.1%	597	(516 - 679)	1.59	(0.71 - 3.54)	0.257
Severe	4	7	36.4%	527	(378 - 676)	2.40	(0.81 - 7.11)	0.115
Institution								
JRH	37	133	21.8%	618	(585 - 651)			
UCLH	7	13	35.0%	575	(472 - 679)	1.71	(0.76 - 3.85)	0.191
Subsite (Combined)								
Oral Tongue	19	74	20.4%	628	(586 - 671)			
Lower OC	17	54	23.9%	605	(551 - 659)	1.27	(0.66 - 2.45)	0.474
Upper OC	8	18	30.8%	587	(502 - 672)	1.47	(0.64 - 3.36)	0.359
Location								
Left	26	65	28.6%	593	(545 - 640)			
Midline	3	7	30.0%	573	(424 - 721)	1.18	(0.36 - 3.89)	0.790
Right	15	74	16.9%	640	(598 - 683)	0.59	(0.31 - 1.11)	0.099
cT								
1	11	52	17.5%	651	(607 - 696)			
2	13	50	20.6%	614	(557 - 671)	1.24	(0.55 - 2.76)	0.604
3	2	7	22.2%	599	(438 - 760)	1.23	(0.27 - 5.54)	0.790
4	18	37	32.7%	570	(508 - 633)	1.97	(0.93 - 4.16)	0.078
cN								
0	21	111	15.9%	658	(628 - 687)			
1	21	34	38.2%	525	(453 - 596)	2.86	(1.56 - 5.24)	<0.001
2	2	1	66.7%	186	(153 - 218)	15.67	(3.45 - 71.19)	<0.001

	E	C	E%	Mean	95% CI	HR	95% CI	p
cStage								
1	7	50	12.3%	682	(645 - 718)			
2	6	40	13.0%	661	(609 - 713)	1.06	(0.36 - 3.16)	0.917
3	13	19	40.6%	498	(399 - 597)	4.17	(1.66 - 10.47)	0.002
4	18	37	32.7%	570	(508 - 633)	2.88	(1.20 - 6.90)	0.018
Growth Pattern								
Endophytic	26	79	24.8%	596	(551 - 642)			
Exophytic	10	32	23.8%	621	(559 - 683)	0.87	(0.42 - 1.81)	0.716
MRI CLNM								
No	27	118	18.6%	640	(608 - 672)			
Yes	12	20	37.5%	542	(454 - 631)	2.22	(1.13 - 4.39)	0.021
MRI ECS								
No	36	138	20.7%	629	(599 - 660)			
Yes	3	1	75.0%	343	(107 - 578)	5.46	(1.67 - 17.78)	0.005
Midline								
Uninvolved	31	123	20.1%	627	(593 - 661)			
Abuts	9	20	31.0%	578	(493 - 663)	1.59	(0.76 - 3.33)	0.222
Crosses	4	3	57.1%	480	(297 - 662)	3.14	(1.11 - 8.90)	0.031
Imaging Bone								
No	28	111	20.1%	630	(595 - 665)			
Yes	15	35	30.0%	575	(508 - 641)	1.54	(0.82 - 2.89)	0.176
pT								
1	13	64	16.9%	651	(609 - 693)			
2	10	36	21.7%	625	(566 - 685)	1.37	(0.60 - 3.12)	0.455
3	3	7	30.0%	567	(408 - 725)	1.88	(0.54 - 6.6)	0.324
4	18	38	32.1%	558	(492 - 624)	2.16	(1.06 - 4.42)	0.034
pN								
0	15	111	11.9%	679	(653 - 705)			
1	9	10	47.4%	503	(392 - 615)	4.84	(2.12 - 11.09)	<0.001
2	20	23	46.5%	468	(384 - 552)	5.55	(2.83 - 10.86)	<0.001
pStage								
1	8	57	12.3%	680	(644 - 716)			
2	3	24	11.1%	684	(632 - 737)	0.93	(0.25 - 3.52)	0.919
3	6	12	33.3%	577	(472 - 682)	2.92	(1.01 - 8.41)	0.047
4	27	51	34.6%	541	(483 - 599)	3.43	(1.56 - 7.55)	0.002
Grade								
Well	5	33	13.2%	674	(624 - 724)			
Moderate	28	97	22.4%	618	(580 - 655)	1.82	(0.70 - 4.71)	0.218
Poor	10	9	52.6%	442	(313 - 570)	5.70	(1.94 - 16.68)	0.002
Invasive Front								
Cohesive	5	35	12.5%	679	(636 - 722)			
Non-Cohesive	36	98	26.9%	591	(550 - 631)	2.55	(1.00 - 6.51)	0.050

	E	C	E%	Mean	95% CI	HR	95% CI	p
Surgical Margin								
Clear	16	71	18.4%	635	(593 - 678)			
Close	21	58	26.6%	600	(550 - 651)	1.48	(0.77 - 2.83)	0.239
Involved	5	16	23.8%	612	(518 - 705)	1.29	(0.47 - 3.52)	0.619
PVI								
No	28	122	18.7%	641	(609 - 672)			
Yes	16	22	42.1%	490	(402 - 577)	3.09	(1.66 - 5.73)	<0.001
PNI								
No	26	98	21.0%	627	(591 - 664)			
Yes	18	46	28.1%	580	(520 - 641)	1.53	(0.84 - 2.79)	0.168
Pathology Bone								
No	32	114	21.9%	622	(587 - 657)			
Yes	12	31	27.9%	579	(507 - 652)	1.38	(0.71 - 2.69)	0.339
Pathology ECS								
No	26	121	17.7%	647	(617 - 678)			
Yes	18	25	41.9%	502	(420 - 584)	2.89	(1.58 - 5.29)	0.001
Positive CLNM								
No	15	111	11.9%	679	(653 - 705)			
Yes	29	33	46.8%	478	(410 - 546)	5.31	(2.84 - 9.92)	<0.001
Type of CLNM								
None	15	111	11.9%	679	(653 - 705)			
Occult	8	14	36.4%	555	(460 - 651)	3.53	(1.49 - 8.36)	0.004
Clinical	21	20	51.2%	437	(350 - 525)	6.42	(3.30 - 12.49)	<0.001
MRI Volume (median)								
<= 3.0 cm^3	16	81	16.5%	650	(612 - 688)			
> 3.0 cm^3	28	65	30.1%	577	(528 - 626)	1.91	(1.03 - 3.52)	0.040
MRI Volume (quartiles)								
<= 0.5 cm^3	7	44	86.3%	673	(629 - 716)			
> 0.5, <= 8.0	19	75	79.8%	622	(578 - 666)	1.59	(0.67 - 3.79)	0.292
> 8.0 cm^3	18	27	60.0%	528	(454 - 603)	3.52	(1.47 - 8.43)	0.005
MRI Thickness (median)								
<= 11.0 mm	16	78	17.0%	646	(607 - 686)			
> 11.0 mm	28	68	29.2%	582	(535 - 630)	1.79	(0.97 - 3.32)	0.062
MRI Thickness (quartiles)								
<= 4.0 mm	4	39	90.7%	694	(655 - 733)			
> 4.0, <= 16.0	16	73	82.0%	635	(592 - 678)	2.09	(0.7 - 6.25)	0.188
> 16.0 mm	24	34	58.6%	529	(464 - 594)	5.01	(1.74 - 14.43)	0.003
MRI Diameter (median)								
<= 18.5 mm	9	54	14.3%	670	(630 - 710)			
> 18.5 mm	35	92	38.0%	586	(544 - 627)	2.15	(1.03 - 4.47)	0.041

	E	C	E%	Mean	95% CI	HR	95% CI	p
MRI Diameter (quartiles)								
<= 0.0 mm	4	32	88.9%	687	(641 - 733)			
> 0.0, <= 18.5	15	55	78.6%	619	(568 - 671)	2.07	(0.69 - 6.25)	0.195
> 28.5 mm	25	59	70.2%	579	(528 - 630)	2.95	(1.03 - 8.48)	0.045
Pathology Diameter (median)								
<= 15.5 mm	10	50	16.7%	660	(617 - 703)			
> 15.5 mm	32	89	26.4%	590	(548 - 633)	1.79	(0.88 - 3.63)	0.110
Pathology Diameter (quartiles)								
<= 10.0 mm	6	27	81.8%	650	(586 - 714)			
> 10.0, <= 25.0	15	66	81.5%	642	(600 - 684)	1.01	(0.39 - 2.61)	0.977
> 25.0 mm	21	47	69.1%	564	(504 - 624)	1.94	(0.78 - 4.8)	0.153
Pathology Depth (median)								
<= 6.0 mm	11	63	14.9%	659	(618 - 700)			
> 6.0 mm	32	78	29.1%	580	(535 - 625)	2.08	(1.05 - 4.12)	0.036
Pathology Depth (quartiles)								
<= 3.0 mm	4	28	87.5%	690	(646 - 734)			
> 3.0, <= 10.0	15	64	81.0%	624	(575 - 673)	1.84	(0.61 - 5.54)	0.280
> 10.0 mm	24	50	67.6%	569	(513 - 624)	2.97	(1.03 - 8.55)	0.044

Continuous Variables

	Mean	95% CI	Median	IQR	HR	95% CI	p
Age							
Censored	60.5	(58.1 - 62.9)	50.3	(60.8 - 70.9)	1.01	(0.99 - 1.03)	0.240
Event	62.9	(58.6 - 67.2)	56.8	(62.1 - 74.7)			
Pathology Diameter							
Censored	23.3	(20.9 - 25.8)	12.3	(20.0 - 32.0)	1.02	(1.00 - 1.04)	0.011
Event	29.3	(24.1 - 34.6)	16.0	(30.0 - 42.0)			
Pathology Depth							
Censored	9.6	(8.4 - 10.9)	4.0	(8.0 - 13.5)	1.04	(1.00 - 1.08)	0.044
Event	12.5	(9.8 - 15.1)	6.0	(11.0 - 18.0)			
Pathology Diameter ($log_{10}X$)							
Censored	1.30	(1.21 - 1.32)	1.10	(1.30 - 1.51)	1.83	(0.79 - 4.28)	0.161
Event	1.40	(1.23 - 1.48)	1.20	(1.48 - 1.62)			
Pathology Depth ($log_{10}X$)							
Censored	0.90	(0.79 - 0.91)	0.60	(0.90 - 1.13)	2.02	(1.05 - 3.89)	0.035
Event	1.00	(0.85 - 1.10)	0.80	(1.04 - 1.26)			
MRI Volume							
Censored	5.4	(3.7 - 7.0)	0.3	(2.2 - 6.2)	1.03	(1.01 - 1.04)	0.001
Event	11.7	(7.6 - 15.8)	1.2	(5.3 - 20.8)			
MRI Volume ($log_{10}X$)							
Censored	0.50	(0.46 - 0.61)	0.10	(0.51 - 0.86)	2.81	(1.57 - 5.05)	<0.001
Event	0.80	(0.67 - 0.99)	0.30	(0.80 - 1.34)			
MRI Thickness							
Censored	10.5	(9.1 - 11.9)	3.7	(10.5 - 15.4)	1.04	(1.02 - 1.07)	0.002
Event	16.0	(12.8 - 19.3)	7.4	(15.1 - 24.9)			
MRI Thickness ($log_{10}X$)							
Censored	0.90	(0.78 - 0.94)	0.70	(1.06 - 1.21)	2.67	(1.24 - 5.76)	0.012
Event	1.10	(0.96 - 1.22)	0.90	(1.21 - 1.41)			
MRI Diameter							
Censored	24.4	(21.3 - 27.5)	12.8	(23.0 - 35.0)	1.02	(1.01 - 1.03)	0.001
Event	35.7	(29.4 - 42.0)	22.2	(35.9 - 50.1)			
MRI Diameter ($log_{10}X$)							
Censored	1.10	(1.04 - 1.25)	1.10	(1.38 - 1.56)	2.41	(1.19 - 4.90)	0.015
Event	1.40	(1.27 - 1.57)	1.40	(1.57 - 1.71)			