Assessment of spatially inhomogeneous intra-organ radiation dose response in salivary glands

by

Haley Clark

B.Sc. Hon. in Physics, The University of Alberta, 2010

M.Sc. in Physics, The University of British Columbia, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Physics and Astronomy)

The University of British Columbia

(Vancouver)

April 2017

Abstract

Cancers treated with radiotherapy must be adequately irradiated to suppress growth at the site of origin. To achieve doses high enough to attain 'local control' and inhibit growth of metastases, surrounding normal tissues are selectively co-irradiated. Current clinical practice for head-and-neck cancers involves salivary gland irradiation. Threshold doses that minimize adverse induced toxicities are currently based on whole-organ mean dose. Modern radiation delivery techniques are able to sculpt the dose profile to accommodate sub-organ irradiation, but knowledge of the relative importance of sub-organ structures remains unknown. As tissue-sparing techniques improve, assessment of the normal tissue toxicity risk becomes increasingly important.

Loss of salivary function and xerostomia (subjective dry mouth) are common normal tissue toxicities in head-and-neck cancer patients. Radiotherapy-induced dysfunction and xerostomia can drastically reduce oral hygiene and health and may negatively impact the ability to eat, speak, sleep, or swallow. These pervasive toxicities detract from overall quality of life and can be permanent, perpetuating the negative impact.

The purpose of this work is to quantify the relative importance of spatial regions within the major salivary glands for late salivary function (i.e., 'regional effects'). The ultimate aim is to improve knowledge of toxicity risk. Broad regional effects have been noted in rat parotid, and it has recently been claimed that a localized 'critical region' has been located in human parotid glands. Furthermore, a morphological dependence on the dose profile has been noted for subjective xerostomia. Clinical trials involving lobe and region sparing are underway, yet comprehensive quantification of the importance of

sub-organ structures remains unknown.

To this end, the association between radiation dose delivered to regions within the largest salivary glands and measurements of whole-mouth salivary flow is quantified. Independent analysis procedures are developed that are capable of quantifying the relative importance of sub-segments. Evidence is found that sub-segments are inhomogeneously important for maintenance of late salivary flow, with the caudal parotid aspects having greatest importance. An imaging protocol is developed which may help pinpoint specific tissues or functional units residing within these regions.

Preface

The identification and design of this research program was completed collaboratively by Vitali Moiseenko, Steven Thomas, Jonn Wu, Allan Hovan, and myself. In particular, the original idea was established, and the collection of patient data began, prior to Haley Clark's involvement in the project.

The collection of patient data has been ongoing for nearly a decade. Collection was performed by the department of oral oncology at the British Columbia Cancer Agency (BCCA) Vancouver, Fraser Valley, and Centre for the North sites. Patient data was managed primarily by Haley Clark, but the earliest efforts before his involvement were headed by Vitali Moiseenko. Clinical contour verification and quality assurance was performed by Jonn Wu.

Unless otherwise indicated, all methodology and analysis presented herein is original work designed and performed by Haley Clark under the supervision and guidance of Steven Thomas, Stefan Reinsberg, Jonn Wu, and Vitali Moiseenko. Hal Clark, alone, was responsible for writing this document.

This work is a continuation of, and expansion upon, earlier work by Haley Clark which was summarized in a MSc thesis submitted to the University of British Columbia (UBC) in October 2010 entitled '*On the Regional Dose Susceptibility of Parotid Gland Function Loss and Recovery: An Effort Toward Amelioration of Radiotherapy-Induced Xerostomia*' [1]. Though the theme is similar to the present thesis, there is no overlap in the work described. All methodology, analysis, and other developments described in the present thesis were performed after acceptance of the MSc thesis (except where clearly and explicitly described, such as for comparison purposes).

Ethics approval from the UBC BCCA REB was granted for the collection and analysis of patient data under the title *A review and comparison of salivary function toxicity from standard techniques in head-and-neck cancer at BCCA and calculation of a radiation dose-salivary function response curve* and certificates H01-02073 and H07-02073.

The introductory chapters in part I were written for the sole purpose of inclusion in this document. The contents of chapters 4, 6 and 14 have not been submitted to academic journals or scientific conferences, but may be in whole or in part in the future. Some figures in chapter 2 were published in the author's MSc thesis (see [1]) in the UBC cIRcle archives and are adaptations by the author from the work of Toldt and Dalla Rosa [2] (now in the public domain, but originally published by the Rebman company, New York in 1919). Some figures in chapter 4 were published in the *Journal of Physics: Conference Series* (2014; vol. 489, pp. 012009) during the author's MSc thesis under the title '*Automated segmentation and dose-volume analysis with* `DICOMautomaton`.' The list of authors was Haley Clark, Steven Thomas, Vitali Moiseenko, Richard Lee, Bradford Gill, Cheryl Duzenli, and Jonn Wu. Haley Clark was responsible for generation of the figures, writing the manuscript, developing the methods, writing all software, ans performing all analysis.

The contents of chapter 11 have been submitted for publication in an academic journal under the title '*Prefer Nested Segmentation to Compound Segmentation.*' The list of authors was Haley Clark, Stefan Reinsberg, Vitali Moiseenko, Jonn Wu, and Steven Thomas. Haley Clark identified the research problem, developed the study and methodology, performed all analysis, and wrote all drafts.

The contents of chapter 12 have been submitted for publication in an academic journal under the title '*Caudal Aspects of the Parotid Gland are Most Important for Radiation-Induced Salivary Dysfunction.*' The list of authors was Haley Clark, Steven Thomas, Jonn Wu, Allan Hovan, Carrie-Lynne Swift, and Stefan Reinsberg. Haley Clark identified the research problem, developed the study and methodology, performed all analysis, and wrote all drafts.

The contents of chapter 13 have been submitted for publication in an academic journal under the title '*Fine segmentation shows anterior-caudal parotid is most important for salivary loss.*' The list of authors was Haley Clark, Steven Thomas, Stefan Reinsberg, Allan Hovan, Vitali Moiseenko, and Jonn Wu. Haley Clark identified the research problem, developed the study and methodology, performed all analysis, and wrote all drafts.

The contents of chapter 15 are an updated version of an early manuscript that was later published under the title '*Development of a method for functional aspect identification in parotid using dynamic contrast-enhanced magnetic resonance imaging and concurrent stimulation*' in Acta Oncologica (2015 Oct 21; vol. 54, no. 9, pp. 1686-90; Taylor & Francis Ltd., http://www.tandfonline.com/loi/ionc20) by Haley Clark, Vitali Moiseenko, Thomas Rackley, Steven Thomas, Jonn Wu, and Stefan Reinsberg [3]. Haley Clark developed the study and methodology, orchestrated all imaging and data collection, designed experimental apparatus, performed all analysis, and wrote all drafts.

Permission has been sought and granted to reproduce all submitted and published materials, both from publishers and co-authors, where applicable.

Table of Contents

List of Tables

List of Figures

Glossary

AIC	Akaike's Information Criterion
ARL	AIC Relative Likelihood
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Effects
DCE-MRI	Dynamic Contrast-Enhanced Magnetic Resonance Imaging
DICOM	Digital Imaging and Communication in Medicine
DOF	Degrees of Freedom
DVH	Dose-Volume Histogram
EORTC	European Organization for Research and Treatment of Cancer
IMRT	Intensity-Modulated Radiotherapy Treatment
KPS	Karnofsky Performance Status
LENT-SOMA	Late Effects Normal Tissue - Subjective, Objective, Management, Analytic late toxicity grading system
MAE	Mean Absolute Error
MSE	Mean Squared Error
MR	Magnetic Resonance

MNAR	Missing Not At Random
NTCP	Normal Tissue Complication Probability
PCA	Principal Component Analysis
PET	Positron Emission Tomography
OAR	Organ-at-Risk
QCD	Quartile Coefficient of Dispersion
QoL	Quality-of-Life
QUANTEC	Quantitative Analysis of Normal Tissue Effects in the Clinic
REB	Research Ethics Board
RMSE	Root-Mean-Square Error
RSE	Residual Standard Error
RSS	Residual Sum of Squares
ROI	Region of Interest
RTOG	Radiation Therapy Oncology Group
SAR	Specific Absorption Ratio
VMAT	Volumetric Arc Therapy

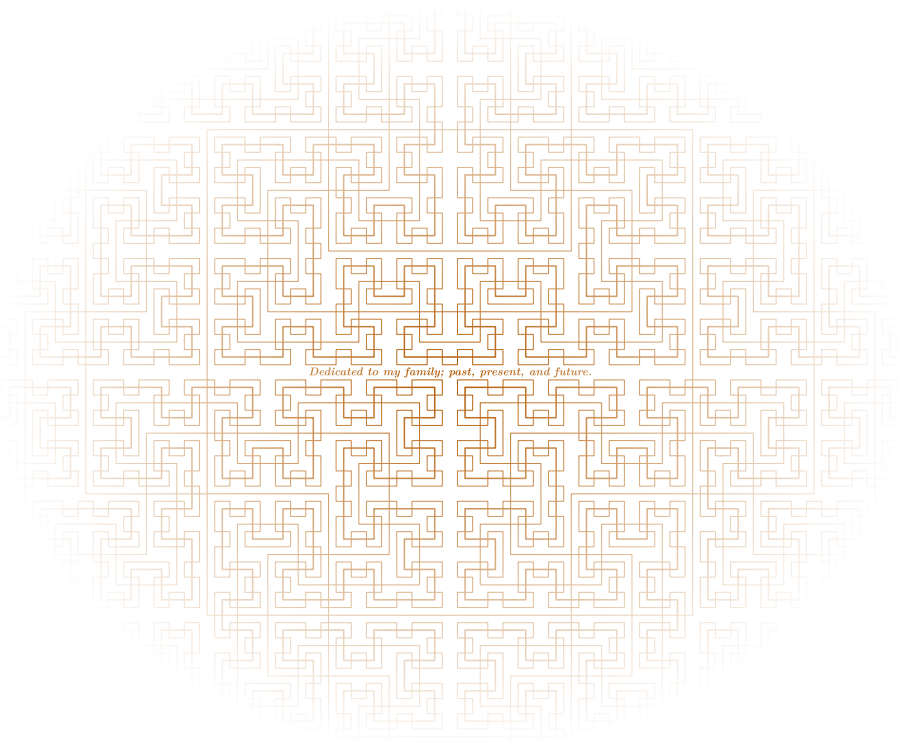
Acknowledgments

The author would like to express gratitude for financial support from the BC Cancer Agency, the University of British Columbia, and the Walter C. Sumner Memorial Foundation.

The author would also like to express gratitude for support from Cheryl Duzenli, Kirpal Kohli, Allan Hovan, Bradford Gill, the BCCA Machine Shop staff, and the UBC MRI staff. All kindly offered their insights or facilitated the author's research efforts.

The author was lucky to have an amicable, flexible, and tolerant committee. The author wishes to thank Steven Thomas, Stefan Reinsberg, Jonn Wu, and Vitali Moiseenko for their support. Each contributed in meaningful ways to the development of this work. Each went the extra mile, always.

Finally, the author relied on love and support from family, especially Sarah.


Dedicated to my family; past, present, and future.

Chapter 1

Introduction

1.1 Motivation and Overview

This thesis concerns the most prevalent radiotherapy-induced toxicities in head-and-neck cancer patients, *salivary gland dysfunction* and *xerostomia* [10]. Dysfunction manifests as (objective) reduction of salivary flow, and xerostomia is the (subjective) condition of persistent dry mouth. These conditions are ultimately distinct. However, in some sense, for example in typical clinical settings, they may be considered two sides of the same coin.

Organs are composed of a heterogeneous mix of tissues. Some types are more radiosensitive than others, and some types are more critical for organ function than others. Currently, it is assumed (clinically) that the tissues that are both radiosensitive and critical for organ function (hereafter referred to as 'critical' tissues) are distributed homogeneously throughout the major salivary glands. If a consistent, inhomogeneous distribution with spatial clustering were observed, it would be known as a 'regional effect.' The converse, 'dose-volume effects,' are evidence for the homogeneous distribution of critical tissues.

Strong dose-volume effects have been known for many decades, so the clinical approximation of a homogeneous distribution has thus-far been constructive. This was partially due to the clumsiness of traditional radiotherapy delivery. Modern technologies, however, have become both precise and accu-

rate. Advancements primarily in computational support technologies have precipitated advancements in machine delivery, improvements in planning capabilities and accuracy, and enabled a plethora of substantial localization improvements through technologies such as on-board imaging [11, 12]. There is correspondingly more control over the delivered dose profile in modern times, and cracks in the homogeneous distribution approximation are beginning to show.

At the same time, overall increases in human longevity and the efficacy of cancer treatments, along with a rise in the number of cancer incidences, have caused the number of cancer survivors to explode [13]. Between 1970 to 2007 Canadian cancer incidence more than *tripled* [14]. Accordingly, patient hardship has become increasingly important and research into methods for improving patient outcomes has burgeoned. A culmination of circumstances involving improvements in treatment technology and the steady erosion of the homogeneous distribution approximation, as well as the growing need to improve patient outcomes have resulted in a clinical need for this thesis – the broad aim of which is to quantify regional effects in the major salivary glands.

More specifically, **the first aim of this thesis is to demonstrate regional effects in the parotid gland,** primarily using parotid gland dysfunction. It is academically interesting to demonstrate such an effect. However, there is also a pressing clinical need to reduce xerostomia and dysfunction incidence and severity. Acknowledging the effect is merely the first step. If the effect can be quantified and incorporated into treatment planning, it may improve toxicity risk estimates and thereby improve patient outcomes. **The second aim of this thesis is therefore to quantify any regional effects.** Preferably in a consistent way rather than merely locating critical sub-structures.

It is not clear how to accomplish the second aim, in particular, and a great deal of effort was therefore spent devising analytical (and physical) machinery to do so. Because the problem is both multi-factorial (i.e., many relevant structures and candidate regions) and multi-faceted (i.e., many endpoints, including stimulated and unstimulated salivary flow, subjective xerostomia

reports), many approaches were considered and tested. A complete description would be inappropriately voluminous. Therefore, only the most promising or successful approaches are reported herein.

1.2 Thesis Organization

The introductory materials span part I. They begin with an overview of factors marginally related to radiotherapy that might still impact regional effects or the ability to detect them. Salivary gland anatomy and physiology are reviewed in chapter 2. A review of imaging techniques and contouring practices are given in chapter 3. A related discussion of salivary gland morphology and topology, and how they are dealt with in later analysis, follows in chapter 4.

The focus then broadens to incorporate radiotherapy-related factors. A discussion of the rationale for irradiating salivary glands is given in chapter 5, and a discussion of the ensuing post-irradiation dysfunction and xerostomia follows in chapter 6. An overview of the instruments employed to assess these toxicities by measuring or estimating endpoint facets is given in chapter 7.

Finally, a variety of short, independent topics of relevance derived from the literature over the last decade are presented in chapter 8. At this point, the reader should have essential understanding of the issues and will be ready to jump into the analysis. Before doing so, the research questions and 'line of attack' are more precisely recapitulated in chapter 9.

Analytical work is contained in part II. Chapters 10 and 11 are largely independent of radiotherapy, but cover topics instrumental for controlling bias in later analyses. Chapters 12 and 13 describe two related but ultimately dissimilar approaches to assessing regional importance within the parotid gland, and largely accomplish the main goals of this thesis by quantifying and characterizing regional effects using parotid gland and dysfunction facets. Chapter 14 is a follow-up that applies the most successful methods developed in chapters 12 and 13 to similarly quantify regional effects using submandibular gland and xerostomia facets. Chapter 15 is the most analytically-disconnected. It describes an Magnetic Resonance (MR) imaging

protocol that was developed to try probe parotid structures important for function within a volunteer cohort. As a volunteer pilot study, there is no possibility of incorporating radiological effects. However it may in the future present a pathway to radiotherapy treatments that are tailored to an individual's anatomy or physiology, perhaps most promisingly when guided by, or used in conjunction with, the population-level regional importance findings from earlier chapters.

Finally, concluding remarks and avenues for future research are discussed in chapter 16.

Part I

Introductory Topics

Chapter 2

Anatomy and Physiology of Salivary Glands

Humans have several salivary glands. The *major* salivary glands comprise parotid, submandibular, and sublingual glands – those responsible for the majority of saliva production. The major glands are characterized by their large size, significant contribution to whole-mouth saliva, remoteness from the oral mucosa, and encapsulation by connective tissue. Comparatively, the *minor* (and *accessory*) salivary glands are smaller, more numerous, and distributed throughout the oral cavity [15]. Embryologically, the major glands develop from the ectoderm, whereas minor glands develop from the mesoderm [16].

Interestingly, human salivary gland anatomy and physiology is similar to that of many other terrestrial species [17]. Some species possess radically modified glandular structures which are, for example, capable of producing venomous saliva in both invertebrates and vertebrates (including mammals) [18]. Owing to the difficulties inherent to studying human tissues, animal models are frequently studied. Rat parotid and sublingual glands are the most common surrogates, rabbit is not uncommon, while mice and drosophila are seen less often [19] [17]. Morphologically, the feline submandibular glands bear a closer resemblance to human submandibulars [20]; despite vitiation, rat studies dominate.

There is considerable variation in shape, size, and location amongst the major salivary glands. It may be surprising, then, that the major glands are somewhat similar in structure, composition, and function. All salivary glands consist of functional *parenchyma* (i.e., the tissue which specializes in secretion of salivary fluids) and supporting *stroma* (i.e., the connective tissue). Parenchyma are connected to the oral cavity through a network of hollow ducts through which salivary fluid is passed. Duct or 'lumen' formation is concurrent with embryogenic development of salivary glands. Branches of solid epithelial tissues that infiltrate the glands are hollowed via apoptosis to form a series of interlinked ducts that permit free passage of saliva [21]. Apoptosis is mediated by members of a family of proteins known as endogenous caspase inhibitors (the only known apoptosis inhibitors known to occur naturally in mammals), including the aptly named *survivin* [22, 23]. Mature ducts are functional shortly after birth. Secretory cells cluster at the terminus of the ducts in roughly spherical groups, which are referred to as 'endpieces' [24].

Endpieces are composed of acinar cells which are specialized to each gland and grouped into small clusters (similar to the gross structure of a raspberry) known as acini. In humans there are *serous* acinar cells, which secrete serous saliva, and *mucous* acinar, which secrete mucous secretions. Serous saliva is mainly water in content [15]. It has digestive properties and is not otherwise used for lubrication [25]. The salivary fluid produced by mucous acinar cells is thicker and is primarily used for lubrication and protection of surfaces [26]. The two types of saliva are created in various proportions in different glands and are ultimately combined to form whole-mouth saliva. In the endpieces, serous cells tend to form spheroidal clusters while mucous cells form into a tubular shape [24].

The ducts have a distinct, regular branching hierarchy. Endpieces secrete directly into *intercalated* ducts, which themselves drain into *striated* ducts, then *excretory* ducts, and finally into the *main excretory* duct which passes into the oral cavity. Duct stroma have low water permeability – considerably less than the parenchyma. Nonetheless, stroma modify the electrolyte composition of fluid passing through the ducts, primarily via removal of sodium

chloride and addition of potassium and bicarbonate [16, 24, 27].

Other tissues present in major salivary glands include: myoepithelial cells, which may help expel various substances from the duct system [16, 28]; plasma cells within the interstitial regions separating glands, which secrete immunoglobulins; blood vessels, which supply the salivary water; parasympathetic and sympathetic autonomic nerves; lymphatic tissues [4]; and a tough, fibrous connective capsule which envelopes each gland. The capsule interposes various septa within the glands which partition the parenchyma into functional groups or 'lobules.' Collective ducts (i.e., intercalated and striated ducts) span lobules and are mostly contained by the septa, whereas excretory ducts pass within the septa, ultimately draining outside the capsule [24].

It is estimated that the major salivary glands supply approximately 60-90% of total saliva secreted [24, 29] [30]. Naturally, as the minor glands secrete less and are more geographically distributed than the major glands (and are thus generally easier to individually spare from incidental radiation damage), they are thought to play a lesser role in radiotherapy-induced xerostomia.

2.1 Parotid Glands

The parotid glands (figs. 2.1 and 2.2) are the largest of the salivary glands. Each parotid is located in the retromandibular fossa (behind the jaw, below either ear). Parotids are bound posteriorly by the sternocleidomastoid muscle and the mastoid process. They lie posterior to, and wrap around (see fig. 2.2), the posterior border of both the mandible angle and ramus [4].

The parotid capsule is composed of a superficial layer of deep cervical fascia. The stylomandibular ligament – the structure separating the inferior portion of the parotid (the 'tail') from the submandibular gland – is formed by the capsule thickening anteriorly and inferiorly. The parotids tend to swell and enlarge under stress. It has been shown that even bulimia can significantly enlarge the parotid [31]. Sensations of pain as the parotid encroaches against the bounds of the capsule are common [32]. Parotids have

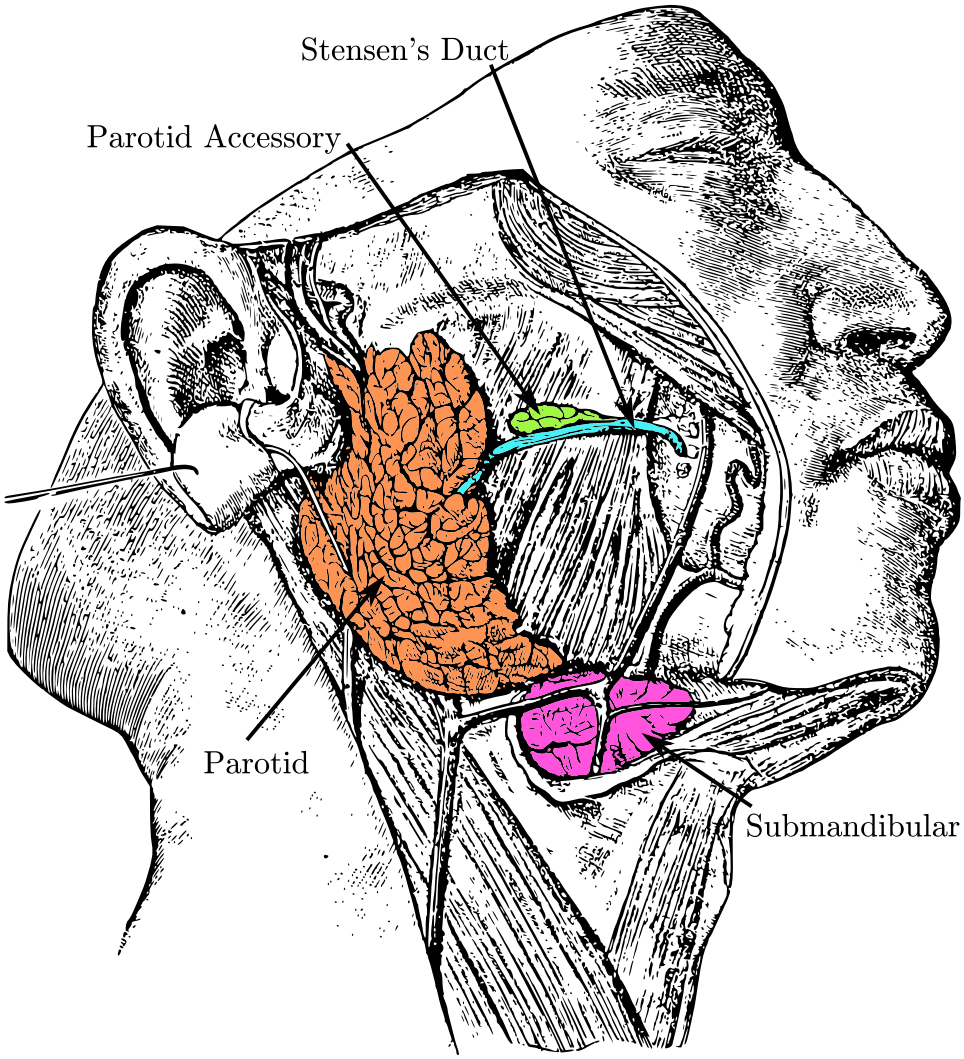


Figure 2.1: Parotid, submandibular, and parotid accessory glands. The ear lobe has been folded so as to not obscure view of the parotid. The parotid accessory gland occurs as a separate gland in 20-40% of the population [4, 5]. Image adapted from Toldt and Dalla Rosa [2], the Rebman company, New York, 1919.

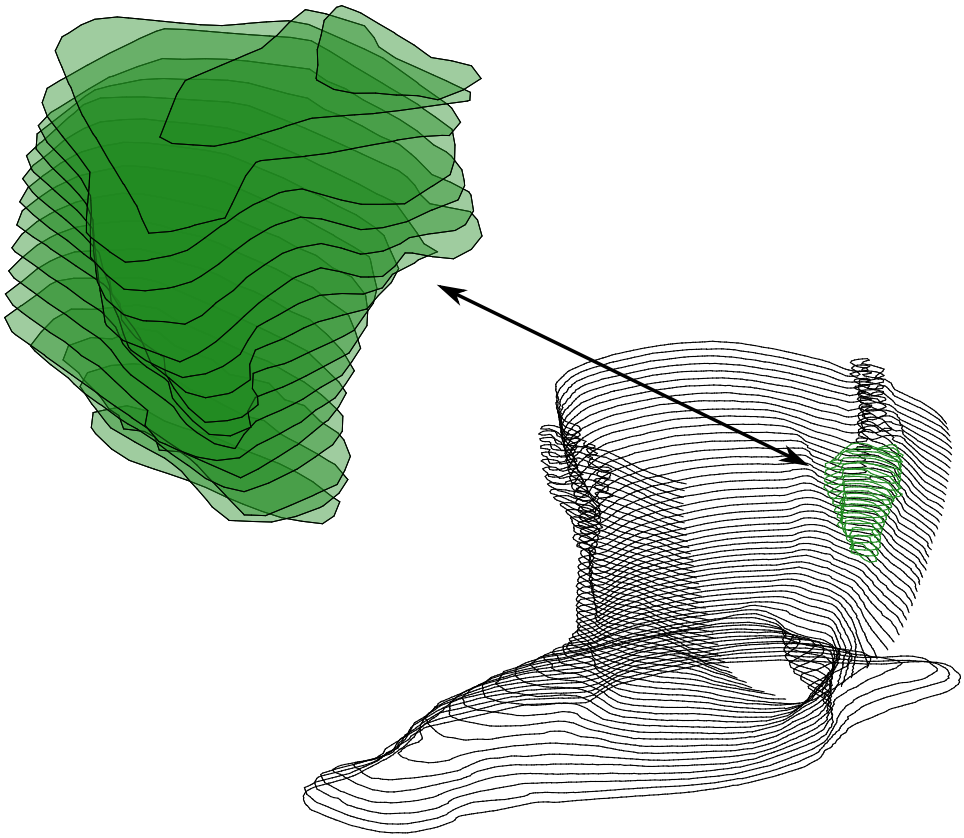


Figure 2.2: Left parotid gland as extracted from computed tomography patient contours demonstrating location, size, and transversely-inverted pyramid shape.

both a superficial lobe (external to the mandible) and a retromandibular lobe (deep to the mandible). The superficial lobe extends superiorly and is substantially larger than the deep lobe. It therefore contains the majority of the glandular tissues. The deep lobe passes within the stylomandibular tunnel (defined as the posterior border of the mandible's ramus, the skull base, and the stylomandibular ligament) [4].

The parotids are supplied with arteries derived from the external carotid, which enters the posteromedial portion of the deep lobe, forks the posterior auricular branch, and terminates into the maxillary and superficial temporal

arteries. The maxillary artery passes out of the parotid along the deep surface, while the superficial temporal artery passes out through the superior pole [33]. Veins follow a similar course. The superficial temporal and maxillary veins join within the parotid to form the posterior facial vein, which ultimately anastomoses with the external jugular vein [4].

Lymphatics pass into and terminate within both superficial and deep lobes. Two to four lymphatic nodes are embedded within the parotid surface [33] while a variable number of nodes are scattered around the vicinity [4]. Nodes drain to the cervical lymph and spinal accessory node groups. Both sympathetic and parasympathetic nerves innervate the parotids. Sympathetics originate from the carotid plexus, whereas parasympathetics originate from the auriculotemporal nerve [4]. Discussion of the role that nerves play in the salivation process is deferred until section 2.5.

Of the two acinar cell types, parotids are comprised almost exclusively of serous type, which produce low-viscosity serous saliva. The parotids mainly secrete under stimulated conditions, such as chewing, when they supply approximately 60% of whole-mouth saliva [15]. They secrete little otherwise [16].

The main excretory salivary duct is 5-7cm long and is known as either the parotid duct or Stensen's duct. It passes over the masseter muscle, pierces the buccinator muscle and buccal fat, and drains into the oral cavity at the upper second molar tooth level [15]. Parotid ducts are not generally visible using conventional Computed Tomography (CT), though they can be imaged using MR, MR sialography, conventional sialography, scintigraphy, ultrasonography, and other less well-known techniques (discussed in chapter 3).

2.2 Submandibular Glands

Following parotids, the largest salivary gland pair are the submandibular glands (figs. 2.1 and 2.3). They are located under the floor of the oral cavity, medial to the mandible [15]. Like parotids, they are characterized as having two major lobes. The larger superficial lobe separates the mandible from the mylohyoid muscle and it itself separated from the parotid by the

aforementioned stylomandibular ligament [4]. The smaller deep lobe lies deep to the mylohyoid muscle and can be easily palpated in the floor of the mouth (see fig. 2.3). Superficial to the submandibular gland are: the platysma muscle, deep cervical fascia, both facial vein and artery, and a portion of the mandibular branch of the facial nerve. Posteriorly the submandibular gland borders the hyoglossus muscle. Both the lingual and hypoglossal nerves trace this border [4].

Both types of acinar cells are present in the submandibular, but the serous usually outnumber the mucous. They mainly secrete saliva under *un*stimulated conditions, during which they supply 70-90% of whole-mouth saliva. During stimulation they produce approximately 20-40% of whole-mouth saliva [15]. The main excretory salivary duct is known either as the submandibular duct or Wharton's duct. Like parotid ducts, submandibular ducts are ∼5cm in length [34]. Exiting from the anterior aspect, submandibular ducts pass anteriorly and superiorly between the hyoglossus and mylohyoid muscles and drains near the lingual frenula, posterior to the lower incisors [15] (see fig. 2.3).

Submandibular glands, like parotids, receive a rich sympathetic innervation from external carotid plexuses. Parasympathetic innervation comes from both facial and glossopharyngeal nerves [4]. The lingual nerve, which supplies sensory innervation to the tongue and carries some parasympathetic fibres, passes over submandibular ducts in two places but directly supplies neither ducts nor submandibulars itself; rather, parasympathetic secretomotor innervation is received from postganglionic fibers that pass through the submandibular ganglion. Lymphatic fluid in the submandibular gland and surrounding tissue drains into the submandibular lymph nodes, which are embedded on the surface of the glands. There are three to six such nodes; small nodes can occasionally be found within the glands.

2.3 Sublingual Glands

The sublingual glands (fig. 2.3) are the smallest and last of the major salivary glands. They are located in the floor of the oral cavity, inhabiting the space

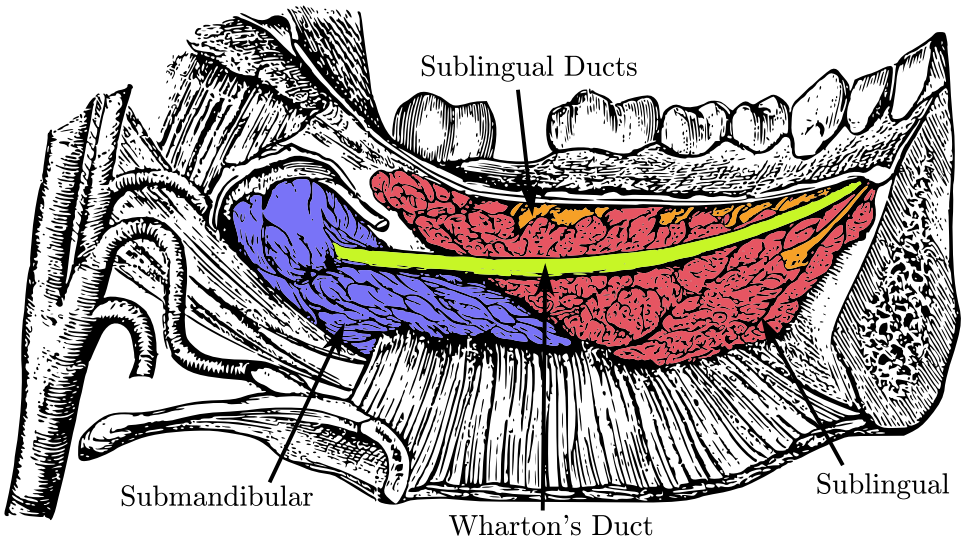


Figure 2.3: View of salivary glands medial to the mandible. Submandibular and sublingual glands can be seen, along with sublingual ducts (draining to the oral cavity) and Wharton's duct. Image adapted from Toldt and Dalla Rosa [2], the Rebman company, New York, 1919.

directly inferior to the mucosa, superior to the submandibular glands [15], and between the mandible and genioglossus muscle. Like the parotids and submandibulars they are composed of both serous and mucous acinar cells; they differ in that the primary type are mucous. Sublingual glands thus secrete a viscous, mucosa-lubricating saliva.

Compared with the submandibular, the sublingual contains few sympathetic nerves. This may be surprising as embryologically both submandibular and sublingual glands develop within the same capsule and descend from the same progenitor cell population [4] (i.e., submandibular and sublingual epithelia branch in the same mesenchyme capsule [35]). *Para*sympathetic nervous innervation is similar to the submandibular. Lymphatic fluid in the sublingual gland will drain into the submental nodes and, along with the submandibulars, into the submandibular nodes.

The duct structure of the sublingual is also similar to the submandibular. Internal ducts drain into the duct of Rivinus, Bartholin's duct, or directly

into the oral cavity through 8-20 small excretory ducts. These small ducts drain under the tongue on the sublingual fold [34] [24].

Upon stimulation the sublinguals produce an estimated 2-5% of whole-mouth saliva [15]. Despite the modest contribution, they, along with the submandibular glands, synthesize the majority of mucin present in saliva [36].

2.4 Minor Glands

The minor salivary glands comprise some 500-1000 small ($\sim$ 1mm in diameter) glands that are distributed throughout the oral cavity and upper aerodigestive tract [24, 37]. They appear with varying probability in the oral mucosa of the hard palate, tongue, within the discontinuities of the mylohyoid muscle [38], the medial border of the cheeks and lips, and throughout the oropharynx [24]. Infrequently, minor gland tissues appear in intraparotid spaces along the neck, bordering the mandible, and within both the middle ear and external ear canal [39].

Despite their multitudinousness, minor glands produce less than 10% of the total mucins present in whole-mouth saliva [15] and contribute less than 5% of whole-mouth saliva. The structure and function of the minor glands differs amongst themselves and from the major glands [24]. For example, minor glands may or may not have an exclusive excretory duct. While minor glands are surrounded by connective tissue, they are not necessarily encapsulated by it as the major glands are. It is generally believed that they are of lesser consequence for total salivary function, though research has shown that they are not completely insignificant [40].

Other types of glands exist (e.g., Von Ebner's glands) but are not discussed here. While such glands are usually of minor importance, they are confusingly not generally classified as minor glands. Accessory glands (e.g., the parotid's accessory glands – see fig. 2.1) are distinct from the parotid in only 20% of the population [4]. For this reason, accessory glands are usually considered to be lobes of major glands and not minor glands *in and of themselves*.

2.5 Saliva

Saliva is composed mainly of water (99.5% by volume) and small amounts of proteins, peptides, and enzymes such as α-amylase, inorganic salts, mucins, bicarbonate, and other compounds (0.5% altogether) [41]. Total salivary flow (stimulated and unstimulated) is estimated to be 1.0-1.5L per day [34] – roughly half comes from the parotids and half from the submandibulars over the span of a day.

Saliva is responsible for moistening and softening food, breaking down starches and a small number of triglyceride lipids, protecting oral mucosa and teeth, and various antibacterial functions. Given the warm, moist environment the mouth presents for aerobic and anaerobic microorganisms, the antibacterial functions of saliva are eminently important for general health. Patients under heavy sedation (which often induces temporary reduction in salivary function) for two weeks or longer have demonstrated a shift in oral microflora from gram-negative to gram-positive species. The normal salivary 'flush' of oral bacteria into the gut in patients with salivary dysfunction thus permits opportunistic pathogens to invade and spread rapidly into the gut and respiratory tract [42, 43]. Salivary enzymes responsible for digestion are inactivated by gastric acidity, so saliva plays a limited role in digestion [24]. However, many proteins thought to be synthesized in salivary endpieces have been detected throughout the body [44]. The function of many proteins and peptides present in saliva are currently unclear [45]. Mucins act as mucosal lubricants; their presence on the mucous membrane surfaces help to maintain a hydrated state [36].

The composition of saliva will change in response to specific stimuli. Human parotid saliva produced in response to citric acid contains far less secretory immunoglobulin-A than that produced in response to mastication [46]. Likewise, stimulation with sweet-tasting solutions elicits a higher protein concentration compared to acidic solutions [47]. In rabbits α-amylase is secreted in higher concentration when fed carrots, but at a threefold-lower volume, compared with pellets [48]. Psychosocial stress is known to strongly affect α-amylase concentrations in human secretions – so much so that it has

been suggested as a reliable, noninvasive indicator of psychosocial stress [49].

The composition of saliva is also known to closely follow circadian rhythms. Similarly, the rate of *un*stimulated whole-mouth salivary output varies with the time of day; low production occurs during sleep. Peak stimulated whole-mouth salivary output occurs daily around 17:00, is lowest at 05:00, and approximately follows a sine wave with a period of one day [25]. Gustatory and olfactory perceptions (i.e., the sensation of taste or smell) or stimulation of either mechanoreceptors (pressure) or nocireceptors (pain) via mastication will greatly increase the rate of salivation – in some cases, up to ten times the baseline [16].

In all salivary glands, production of saliva is stimulated by both parasympathetic and, to a lesser degree, sympathetic nervous systems. The secretion of proteins and salivary fluid is controlled by autonomic nerves [16]. Secretion of saliva occurs when cholinergic parasympathetic nerves emit acetylcholine. Acinar endpieces are evoked and secrete saliva when the acetylcholine binds to the cell's muscarinic receptors. Salivation is a rapid response to nervous stimulation. Cessation of parenchymal secretion occurs rapidly when the parasympathetic nerve stimulation is interrupted [16], but salivation may persist after the stimulus subsides due to residual drainage in the ducts. A few minor salivary glands will spontaneously secrete in absence of nervous stimuli, but maintenance of a continuous, normal secretory rate requires an autonomic nerve supply [50].

2.6 Conclusions

Parotid and submandibular glands contribute the most to whole-mouth saliva. They appear to have complimentary functions, with parotids responsible for the majority of stimulated saliva and submandibulars responsible for the majority of resting/unstimulated saliva. Salivary gland dysfunction leads to increased susceptibility to opportunistic pathogens, which, absent the protective aspects of saliva, are able to pass from the aerodigestive tract into the gut. Both parotids and submandibulars are known to have lymph node involvement, which, as will be discussed in chapter 5, has ramifications for head-and-neck cancer patients treated with radiotherapy.

Chapter 3

Salivary Gland Imaging and Contouring

3.1 Survey of Imaging Techniques

Salivary glands are highly accessible from an imaging perspective. The major parotid gland duct (Stensen's duct) papilla is visible at the upper second molar tooth level and the major submandibular gland duct (Wharton's duct) is visible in the floor of the mouth; both are readily cannulated [15]. Historically, sialographic radiology exploited this accessibility to visualize ducts by flushing a contrast-enhancing agent via cannula [51]. The process is simple; a scout is taken – normally at an oblique lateral-anterior angle, local anesthesia is applied around the papilla of the duct or lidocaine is initially passed through the orifice, the duct is dialated using a lacrimal probe and saline, a contrast agent is flushed, and one or more radiographs are taken to observe agent perfusion [52, 53]. This technique was first recorded in vivo in 1925 and has been employed for nearly a century [54, 55]. The earliest contrast agents contained bismuth, were wholly mercury, or used early precursors to modern iodinated compounds, such as potassium iodide or lipiodol [52, 54, 56, 57]. Intraoral radiography is sometimes employed [58]. However, conventional radiography without cannulation or contrast enhancement, despite having excellent spatial resolution is generally insufficient to resolve salivary gland

ductal structures due to the complex head-and-neck anatomy [59].

Despite long-time use of radiological sialography, several well-known detractors persist. Grievously for patients, it requires painful dilation and cannulation of the duct. It can also cause retention of contrast agent and is of limited use when perfusion is impossible [54, 57]. Other minimally-invasive and non-invasive techniques are also amenable owing to the accessibility. Ultrasound sonography is non-invasive and causes no discomfort. In salivary glands it is primarily used for resolving duct calcifications; older studies found calcification detection capabilities similar to sialography [60]. It can also be used to quantify morphological aspects of salivary glands [61]. The resolution of ducts themselves is (or has historically been) generally poor without dilating the ducts (requiring the examination to be somewhat invasive). So it mostly employed to examine stroma and the protective capsule rather than ducts. Sonography appears to be underused considering it is one of the least invasive techniques available; it is thought that sonography, in general, has somewhat been displaced by the wide availability and applicability of general radiology, CT, and MR [62]. Recent technological advancements in beamforming, ultrafast techniques, and low-cost construction methods may eventually lead to better ductal imaging and greater ubiquity [63, 64, 65, 66, 67].

Scintigraphy is frequently used to image salivary action in parotid and submandibular glands by quantifying excretion of a radiotracer directly from the parenchyma [52, 68, 69]. Many tracers have been investigated, but the most popular are ^{11}C-methionine and ^{99m}Tc-pertechnetate. Because scintigraphy can quantify function, it can therefore be applied to detect radiotherapy-induced dysfunction by comparison of pre- and post-treatment images. However, the highest-precision scintigraphic imagers – those for small animals – have a spatial resolution limit of around 2mm and are therefore not able to resolve fine salivary gland internal structure [70]. Additionally, scintigraphy and other radioisotope-based methods are generally unsuitable for continued observation due to limits on patient exposure, and may therefore limit the total number of imaging applications possible [71].

The current most popular imaging methods for applications in radiotherapy, sialadenitis, and sialolithiasis (i.e., obstruction via calcifications) appear

to be CT and the emerging technique of MR sialography [72]. CT is widely available and offers millimetre or sub-millimetre spatial resolution in the head-and-neck [52, 73]. It is widely applied to assess salivary obstruction and dysfunction, and is capable of quantifying radiotherapy-induced morphological changes [74, 75]. MR sialography is a non-invasive modality that uses *in situ* saliva secretions as an endogenous contrast medium [72]. No additional contrast agents are required. It has been shown to be effective for quantifying radiotherapy-induced dysfunction and duct damage [76]. The flexibility in contrast derivation (i.e., endogenous or exogenous agents) and freedom in choice of coil apparatus (i.e., microscopy coils) makes MR sialography a versatile and attractive imaging tool.

Many other imaging methods have recently emerged or are currently under intense development. Examples include: Sialoendoscopy, which involves insertion of a small endoscope into the parotid duct, but requires essentially the same preparations as sialographic techniques and is of limited use for assessing whole-organ flow [77, 78]; MR spectroscopy, which allows spatially-localized measurement of metabolite concentration in tissues [79]; MR diffusion-weighted imaging [80], which can assess characteristics of gland function; Dynamic Contrast-Enhanced Magnetic Resonance Imaging, which uses the flow of a contrast-enhancing agent through the vasculature ('perfusion') to derive information about blood flow from a volumetric image time series (which is employed later in this thesis) [81, 82, 83]; chemical exchange saturation transfer (CEST), which exploits the chemical exchange between free water and mobile exogenous or endogenous agents [84, 85]; multiparametric methods that combine signals from multiple imaging techniques into novel facets (e.g., cell density) [86]; Positron Emission Tomography (PET), in various capacities [69, 86]; and various methods based on electrical impedance [87, 88, 89, 90]. A theme in modern imaging research appears to be development of multi-modality imaging techniques that can be combined to overcome limitations of individual techniques. Issues with such approaches vary, but often are more costly (both money and time), may require higher cumulative patient doses, and often require image registration. Examples include sialoendoscopy coupled with MR sialography [91], conventional MR

coupled with MR sialography [92], and scintigraphy coupled with CT [93].

The most commonly-used modalities for diagnostic imaging are radiography, sonography, CT, and MR, with CT and MR use growing faster than the other modalities [94, 95, 96]. Presently, CT and MR imaging methods are most common for assessing radiotherapy-induced toxicity in salivary gland. CTs are currently needed for tissue density estimation and radiotherapy treatment planning, so CT scanners are thus widely available. A recent development has been reliable estimation of electron density using only MR to improve MR clinical adoption and reduce reliance on CT [97, 98, 99, 100]. Alternatively, there are efforts to combine fully-capable CT and MR imaging units into a single clinical unit [101]. Contrast-enhanced CT and MR are both minimally-invasive. CT is fast and cheap, MR is less common, less cheap, less fast, and presents more contraindications, including any implanted or embedded ferromagnetic materials and implanted medical devices. However, MR is arguably more versatile, permitting more complicated imaging protocols and a larger number of ways in which contrast can be derived. For example, continual scanning is permitted over long time spans, and the contrast mechanism can be switched on-the-fly. X-ray and CT imaging use ionizing radiation and patients therefore receive radiation dose. Doses received from routine head-and-neck clinical scans are comparable to natural background levels and are therefore acceptable for diagnostic purposes, but they impose limits on the duration and contrast possible [102]. MR scans use no ionizing radiation, sparing patients dose, but can breach Specific Absorption Ratio (SAR) limits. SAR limits present lower biological risk than x-ray dose at typical clinical levels. Overall though, CT is cheaper and has an established presence in the clinic. It is therefore unlikely to abate in the near future. Since the aim of this research is not only to characterize regional effects within the parotid, but also to develop an assessment procedure which can incorporate toxicity risk into treatment planning, and treatment planning is currently performed exclusively with CT images, the almost wholesale reliance on clinical CT for analysis of regional effects has been deliberate. An overview of CT as it pertains to clinical practice at the British Columbia Cancer Agency (BCCA) is given below. Additional MR imaging support was also pursued, so it is

also briefly discussed below.

3.2 Computed Tomography Imaging

Computed Tomography emerged commercially in 1972 as a three-dimensional counterpart to two-dimensional radiographic imaging [103]. Many technical aspects had been dealt with theoretically decades prior. For example, the Radon transform was devised in 1917 for the abstract purposes of function reconstruction over non-Euclidean manifolds and higher-dimension spaces [104]. The primary advantage of CT was not merely that images were digital, but rather that cross-sectional 'slices' of tissue could be generated rather than having to direct x-rays orthographically through a large perpendicular section. The mechanism, in brief, is simple but computationally demanding. An x-ray tube emits photons that are directed through the target tissue. X-rays are selectively absorbed with different probabilities in different tissues (related to the photon scattering cross-section). A detector records x-ray intensity on the opposing side of the target. This projection scheme is iterated in several orientations, but the x-ray source and detector are usually rotated in lockstep. Image reconstruction is a field of ongoing research, but a crude, illustrative image can be generated simply. First, connect a line segment between each individual detector position and the source position at each iteration. If the line segments are given a weight (or grayscale value) related to the intensity, then the total accumulation of line segment weight at each location (i.e., density) in the virtual target will be crudely related to the contrast and an image may be recognizable. However, even in the early days of CT this approach would be considered inappropriate [105]. Modern iterative methods that require less irradiation and thus lesser patient dose are now used [106].

Historically, inter-centre routine CT clinical scan parameters varied and were known to be subjectively influenced by perceived contrast and patient weight [107, 108]. Modern scanners automatically control exposure by adjusting tube current using measured patient beam attenuation. This not only improves image consistency, but can also decrease overall patient dose [109, 110, 111].

Radiocontrast agents can be used to improve CT contrast. At typical clinical CT energies, the photoelectric effect dominates the x-ray absorption spectrum. The enhancement mechanism relies on the photoelectric effect x-ray absorption coefficient, (μ), having a stronger dependence on atomic number (Z) than mass-density (ρ), photon energy (E), and atomic mass (m_A) [112]. Specifically,

$$\mu \propto \frac{\rho Z^4}{m_A E^3}. \tag{3.1}$$

An incident x-ray with energy sufficient to displace an atom's K-shell (i.e., inner-most shell) electron will be absorbed. Since this energy – called the 'absorption edge' – increases as Z increases, increasing Z while keeping gross tissue ρ constant will alter the x-ray absorption spectrum. Directed uptake of high-Z radiocontrast agents in soft tissues will therefore improve image contrast. Since such agents alter effective electron densities, which are derived from CT images, and then the densities are used for radiotherapy treatment planning, there is a small discrepancy introduced between planned and delivered dose. The effects are within normal tolerances and work-around techniques have been developed if the impact on dose is too great [113, 114, 115, 116]. Radiocontrast agents are employed for some routine diagnostic imaging at the BCCA; primarily ioversol (Optiray 300, Mallinckrodt Pharmaceuticals), an iodide-based, nephrotropic, low osmolarity agent. Dosage is uniformly 120mL for patients $\geq$45kg. It is injected intravenously and followed by a saline flush.

A generic CT imaging protocol consists of a 'pilot' scan to identify tissue target boundaries, a 'localizer' scan with coarse spatial resolution to align the imaging coordinate system and ensure tissue targets are within the field of view, and a final, single-pass axial scan.

3.2.1 Current Clinical Practices

Routine CT imaging parameters remained fixed for head-and-neck cancer patients at the BCCA over 2005-2015. For a sampling of 886 head-and-neck

cancer patients (that were assigned a study identifier and considered for later analysis), tube characteristics were nominally 120 kVp (median: 120.0 kVp, mean: 120.7 kVp) and 350 mA (median: 350.0 mA, mean: 332.3 mA). The median total patient exposure was 93.0 mAs (20^{th} percentile: 77.0 mAs; 80^{th} percentile: 93.0 mAs). Patients were almost always positioned head-first in the supine position (99.9% of the time). Slice thickness was nominally 2.5mm (91%). CT simulators used to image patients in this cohort included GE Medical Systems LightSpeed RT16 (approximately 83% of scans) or Optima CT580 (8%), a Picker International, Inc. (n.b. acquired by Philips Healthcare) PQ 5000 (7%), and a Philips Healthcare Brilliance Big Bore (2%).

Example routine CT images with common imaging characteristics (at 120 kVp and 350 mA) demonstrating Region of Interest (ROI) contours and tissue contrast differences are shown in fig. 3.1. Images showing visible internal parotid sub-structure and demonstrating contrast enhancement (same patient and scan as fig. 3.1) are shown in fig. 3.2.

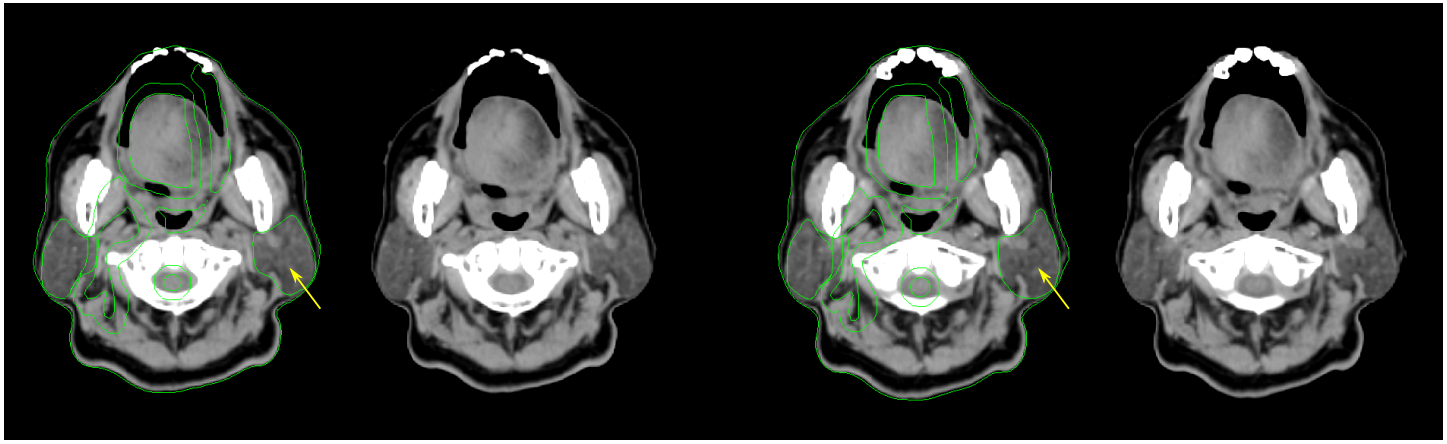


Figure 3.1: Two example BCCA routine axial CT images (at 120 kVp and 350 mA) showing parotid contours around 2cm inferior to the ear canal. The left parotid is indicated in each, and a view with and without ROI is shown to demonstrate tissue contrast differences. Other ROI include the pharynx, spinal cord and margin, clinical and planning target volumes encompassing the tongue and right nodes, and a portion of the left oral cavity that has been subtracted from the target volume for sparing purposes.

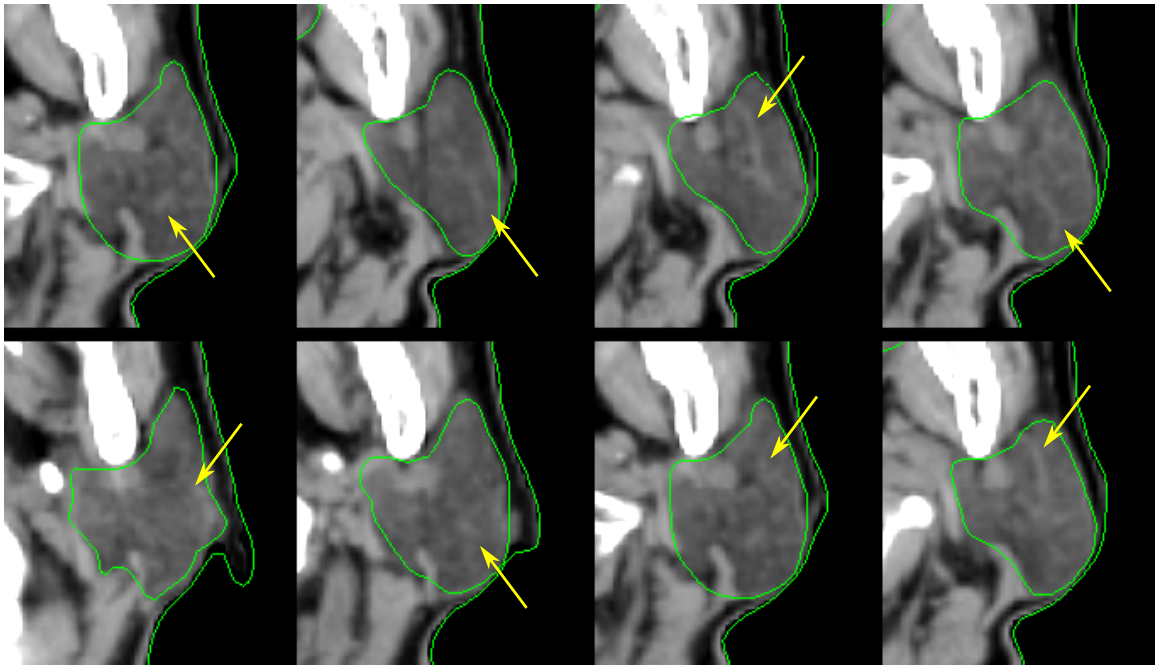


Figure 3.2: Example BCCA routine axial CT images (at 120 kVp and 350 mA) at various levels demonstrating visibility of sub-structures within the left parotid (variously indicated; all are most likely vasculature owing to the relatively low permeability of acinar cells to ioversol).

3.3 Magnetic Resonance Imaging

Clinical whole-body MR imaging emerged in the late 1970's and early 1980's [117]. The first published images were in 1973-74 with the proposed name 'magnetic resonance zeugmatography' [118, 119]. The name is now disused, but the technique was quickly adopted by the community. The first clinical scanners were available in the early 1980's and were generally <1T, though 1.5T scanners followed shortly [117].

Contrast in MR imaging ultimately derives from the excitation and subsequent relaxation of particles that develop a magnetic dipole moment when placed in a magnetic field. The only requirement is that the particles possess spin, so elementary particles such as electrons and neutrinos, composite particles such as neutrons and protons, and composite objects such as nuclei, atoms, and some molecules could in principle be used for MR imaging. However, the hydrogen atom proton is used exclusively for clinical imaging. In brief, the potential energy of spins within a magnetic field become biased. Collections of particles that develop a magnetic dipole moment will thus become slightly polarized when placed in a static field. Radiofrequency pulses can be used to transition individual spins from state to state. The potential energy for interaction with a static magnetic field ($\vec{B_0}$) and a spin-$\frac{1}{2}$ particle with dipole moment $\vec{\mu}$ is $V = -\vec{\mu} \cdot \vec{B_0}$; the ground spinor state is $|\downarrow\rangle$ with energy $-|\vec{\mu}| \left|\vec{B_0}\right|$ and the excited state is $|\uparrow\rangle$ with energy $+|\vec{\mu}| \left|\vec{B_0}\right|$, so the potential difference between spinor states is $2|\vec{\mu}| \left|\vec{B_0}\right|$. This means the radiofrequency pulse must have a frequency of $2|\vec{\mu}| \left|\vec{B_0}\right| / (2\pi\hbar)$ where $\hbar$ is the reduced Planck's constant. This is the so-called 'Larmor frequency.' For protons the Larmor frequency scales with field strength (in Tesla) like $\sim 42.578 \frac{|\vec{B_0}|}{1T}$MHz which is $\sim$63.867MHz at 1.5T and $\sim$127.734MHz at 3T. After spins have transitioned to the higher energy state and the radiofrequency signal is discontinued, spins will gradually return to the equilibrium polarization through thermal relaxation. Spins transitioning back to the ground state emit a characteristic radiofrequency pulse at the Larmor frequency, which can be detected.

A magnetic resonance imager consists of three essential components: a

static magnetic field ($\vec{B_0}$), an excitation radio, and a measurement radio with appropriate antennae. Additional components are needed to process the data, though rudimentary techniques may yield a reasonable image. Only modest field strengths are needed, for example the Earth's 'ultra low' magnetic field will suffice for some magnetic resonance spectroscopic applications [120]. A table-top scanner can be scavenged for as little as $200-300 [121]. Modern scanners vary in cost considerably, but in the United Kingdom in 2011 a typical 1.5T scanner cost $\$1.4M$USD to purchase and roughly the same in maintenance costs for the lifetime of the scanner [122].

Modern clinical scanners with active magnets generally have field strengths 1.5-3.0T. Manufacturing high field-strength devices is much harder than low field-strength counterparts. The high $|\vec{B_0}|$ means special considerations must be given to all components, internal or external. Homogeneity of the Bo field is important for minimizing image artifacts, and characteristic times depend strongly on $|\vec{B_0}|$ [123]. Superconducting magnets are often used, which require special precautions such as persistent staffing or monitoring, ventilation in case the magnet 'quenches,' and shielding from outside noise sources and fringe magnetic fields. Scanners impose other constraints, including economic and special containment room construction. Older 1.5T scanners required 50-90 tonnes of iron to shield fringe fields [124]. More recent 3.0T scanners themselves weighed $\sim$10 tonnes (even with active shielding, which sheds some bulk). A 7.0T scanner weighs $\sim$30 tonnes and may require 100 tonnes or more of shielding to contain fringe fields [125].

3.3.1 Current Clinical Practices

MR imaging is not used for routine diagnostic imaging for head-and-neck cancer patients at the BCCA. However, as will be discussed in chapter 15, MR imaging of a volunteer cohort was performed. Example images of general tissue differentiation in the head-and-neck are shown in fig. 3.3. Parotids in the same patient at various levels are shown in fig. 3.4 to demonstrate the clear borders between tissues. Both fig. 3.3 and fig. 3.4 are T_1-weighted spin-echo sequences that derive contrast from spin-lattice interaction relaxation ($T_E = 16\text{ms}$, $T_R = 619\text{ms}$).

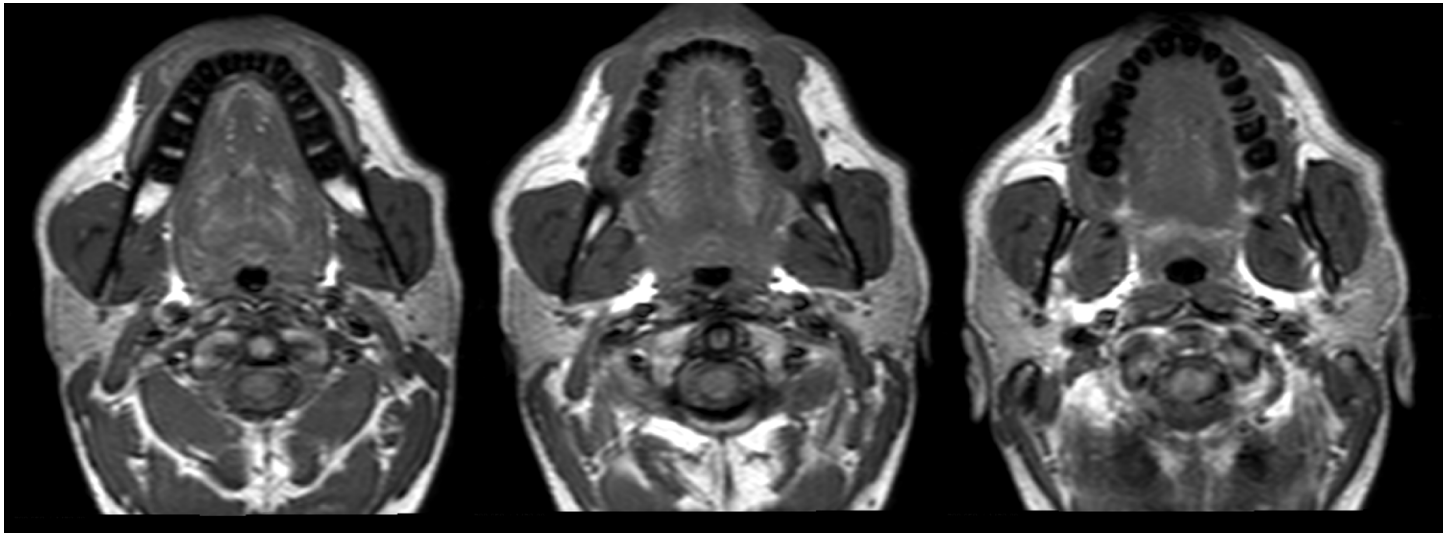


Figure 3.3: Example anatomical (T_1-weighted, with $T_E = 16$ms and $T_R = 619$ms) MR axial images in the vicinity of the ear canal. Tissue differentiation can generally improve upon the equivalent CT scan by adjusting the contrast sensitivity (cf. fig. 3.1). No contrast enhancement agents were administered.

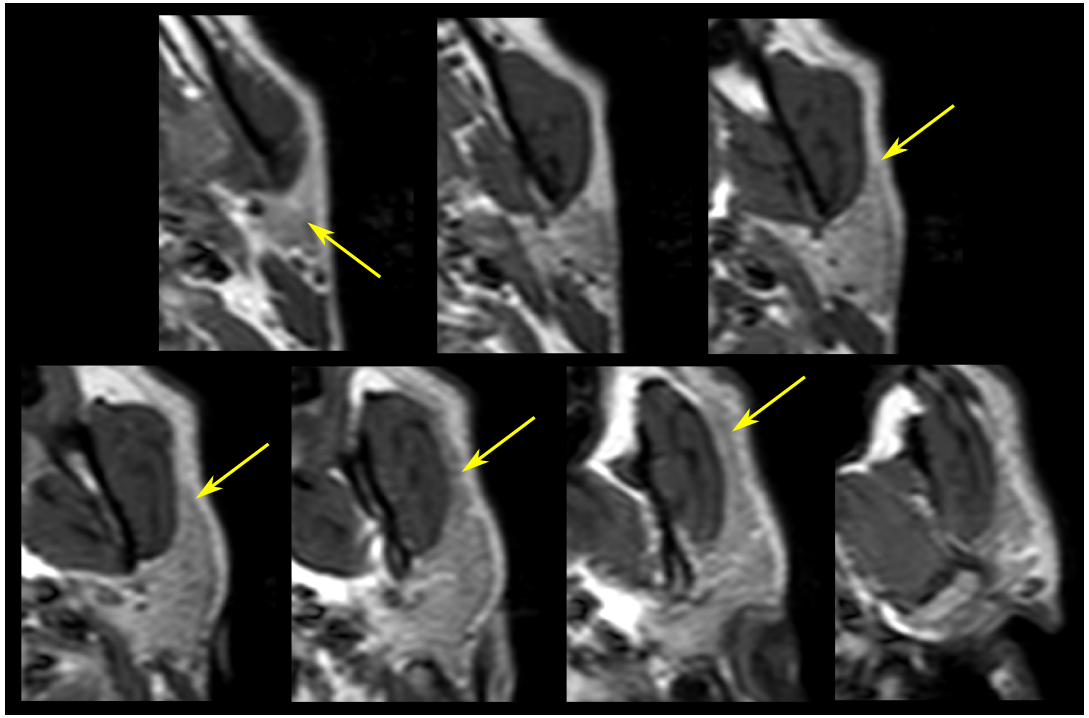


Figure 3.4: Example anatomical (T_1-weighted, with $T_E = 16$ms and $T_R = 619$ms) MR axial images at various levels demonstrating clear tissue borders. No contrast enhancement agents were administered.

3.4 Contouring Practices in the Clinic

One of the key purposes of CT imaging for radiotherapy patients is to create treatment plans. Creation of plans requires two necessary tasks. First, demarcating ROI contours for radiotherapy targets (i.e., diseased tissue) and Organs-at-Risk (OARs), such as parotids, which in some sense can be thought of as 'spectator' tissues that should not be substantially irradiated. The second task is to determine how to deliver the prescribed radiation dose to the targets while minimizing dose to OARs. The second task is simulating the interaction between tissue and radiation and optimization of radiation delivery to achieve several criteria simultaneously (e.g., 70Gy to the primary disease site, dose limits for OARs, dose shape limits in some cases). Due to use of machine optimization it is not always intuitive how small changes to the inputs or optimization criteria will impact candidate plans, and optimization is something of a 'black-box.' It is therefore important to be consistent in contouring practices. Clinical contouring guidelines at the BCCA are fixed for key OARs to help reduce subjectivity.

There are BCCA guidelines for contouring parotid and submandibular glands, the gross oral cavity, laryngopharynx, mandible, lips, brainstem, and optic chiasm and nerves. Not all structures are contoured in every case. For example, some less critical OARs may be omitted for low-dosage plans. Clinical borders for key structures are described in table 3.1.

Parotid contouring instructions are to begin superiorly, identify the parotid gland behind the ramus of mandible, contour every second slice until the inferior border, and then interpolate contours. Submandibular instructions are to begin inferiorly around the level of the carotid bifurcation, identify the submandibular gland lateral to the hyoid bone, contour every second slice until the superior border, and then interpolate contours. The oral cavity is begun superiorly where the hard palate is seen covered with mucosa, every second slice is contoured until the inferior border, and contours are interpolated. Lips are excluded from the oral cavity, and the oropharynx is generally not contoured so the communicating border is taken at the posterior level of the epiglottis or Laryngopharynx.

Structure	Border	Structures
Parotid	Anterior	Masseter muscle, Ramus of Mandible
	Posterior	Sternocleidomastoid muscle, Digastrics muscle posterior belly
	Superior	Zygomatic arch
	Inferior	Fascia between Sternocleidomastoid muscle and Mandible
	Medial	Styloid process, Medial Pterygoid muscle
	Lateral	(open)
Submandibular	Anterior	Platysma muscle
	Posterior	Sternocleidomastoid muscle
	Superior	Mandible, Mastoid process
	Inferior	Diagastric muscles, Epiglottis, Laryngopharynx
	Medial	Hyoid bone, Tongue, Oropharynx
	Lateral	Platysma muscle, Mandible
Oral Cavity	Anterior	Teeth, Mandible
	Posterior	Laryngopharynx, (communicates with the Oropharynx)
	Superior	Palate (hard, soft)
	Inferior	Tongue, Mucosa, Geniohyloid and Mylohyloid muscles
	Medial	Teeth, Mandible
	Lateral	Teeth, Mandible

Table 3.1: BCCA head-and-neck ROI clinical contouring: structures bordering major salivary glands and the oral cavity.

3.4.1 Clinical ROI Statistics

Variations in contouring practices have been described across centres [126, 127], inter-observer[1] in the same centre [128, 129, 130, 131], and intra-observer during repeatability tests [132, 133, 134, 135]. Both OAR and target contouring are impacted.

Recent research efforts into creating 'oracles' – algorithms or computer systems that can establish ground truth for certain structures in a consistent way – have been successful [136, 137]. Most approaches are based on deformation and construction of one or more atlases that have been manually contoured by one or more experts [138]. The *Simultaneous Truth and Performance Level Estimation* approach is common [139]. But salivary organ oracles specifically, while steadily improving in the literature, are not yet reliable enough to establish ground truth without manual intervention [138, 140, 141, 142, 143]. Therefore, it is worthwhile to report sample population (descriptive) statistics for later comparison to help assess similarity of contouring practices or highlight specific differences that may translate into significant differences in analysis.

Simple statistics for all available BCCA ROI (886 patients that had been assigned an anonymous identifier; pre-radiotherapy planning CTs) were estimated (see table 3.2). Strictly $\mathbb{R}^2$ volumetric and surface quantity estimators were employed to avoid topological ambiguities and surface reconstruction issues (discussed in section 4.2). Volume was estimated by the 'slab volume,' which is total planar contour area multiplied by the image slice thickness. It may over- or under-estimate the true volume, especially when ROI are highly curved and the (true) surface area is comparable to the total planar area. Slab volume will exactly estimate the true volume for shapes with faces either parallel or perpendicular to the axial plane (i.e., not oblique), such as axes-aligned rectangles and cylinders. Salivary organs are relatively cylindrical, being elongated along the superior-inferior direction, and are generally not small enough for surface effects to dominate. Total contour polygon

[1] An encompassing term that can refer to oncologists, medical physicists, surgeons, clinical therapists, and physician assistants, but rarely students or those with limited clinical experience.

perimeter length is used to convey information about average circumference of individual contours. It can also be used to roughly approximate surface area when image slice thickness is uniform. In any case, the statistics in table 3.2 are meant for direct comparison with other cohorts, and should be compared to data computed using the same methods.

Deviations from lateral symmetry were unremarkable. Parotids were found to have approximately 3.8× the slab volume of submandibulars. Parotids had approximately 2.7× the total perimeter of submandibulars, but only 1.6× as many contours, suggesting submandibulars are more elongated along the superior-inferior direction. Parotids, in comparison, have a more round shape in the medial-lateral and/or anterior-posterior directions. For comparison, the oral cavity has approximately 2.7× the slab volume and 1.4× the perimeter. While larger, it is morphologically more compact compared to both parotid and submandibular glands.

ROI	ROI Count	ROI Count			Slab Volume (mm^3)			Total Perimeter (mm)		
		20^{th}%	median	80^{th}%	20^{th}%	median	80^{th}%	20^{th}%	median	80^{th}%
Left Parotid	641	20	24	27	21893	29042	39005	1723	2159	2632
Right Parotid	632	20	24	27	21844	29898	38515	1736	2185	2629
Both Parotids	1273	20	24	27	21872	29471	38717	1730	2168	2632
Left Submand.	589	12	15	18	5612	7902	10170	627	799	968
Right Submand.	588	12	15	18	5640	7679	10170	613	790	960
Both Submand.'s	1177	12	15	18	5619	7855	10184	621	793	963
Oral Cavity	507	16	19	22	55529	79774	109114	2376	2944	3719

Table 3.2: BCCA head-and-neck ROI contouring practice statistics for salivary glands and the oral cavity (886 patients examined). ROI count refers to the number of the specified ROI present in the cohort. Sample population 20^{th}, 50^{th} (i.e., median), and 80^{th} percentiles are shown. 'Slab volume' refers to total planar area multiplied by the image slice thickness. Lateral symmetry is strong.

ROI	ROI Count	Ant.-Post. (mm)			Med.-Lat. (mm)			Sup.-Inf. (mm)		
		20^{th}%	median	80^{th}%	20^{th}%	median	80^{th}%	20^{th}%	median	80^{th}%
Left Parotid	641	35.4	39.7	45.3	39.7	47.4	58.7	52.5	60.0	67.5
Right Parotid	632	35.3	40.1	46.0	39.7	47.7	60.1	52.0	60.0	67.5
Both Parotids	1273	35.3	39.9	45.6	39.7	47.5	59.8	52.5	60.0	67.5
Left Submand.	589	20.0	22.8	26.1	23.1	27.3	31.9	32.5	37.5	45.0
Right Submand.	588	20.6	23.2	26.5	23.3	27.2	31.1	32.5	37.5	45.0
Both Submand.'s	1177	20.4	23.1	26.3	23.1	27.3	31.5	32.5	37.5	45.0
Oral Cavity	507	47.0	54.7	63.3	59.6	66.0	74.3	38.0	47.5	55.0

Table 3.3: BCCA head-and-neck ROI extreme linear dimensions (i.e., 'caliper width') along orthogonal anatomical directions (886 patients examined). ROI count refers to the number of the specified ROI present in the cohort. Sample population 20^{th}, 50^{th} (i.e., median), and 80^{th} percentiles are shown. Lateral symmetry is strong.

Extreme linear dimensions (i.e., the distance a caliper would measure) along orthogonal anatomical directions are computed in table 3.3 to help further describe organ morphology. Parotids are shortest in the anterior-posterior direction and largest along the superior-inferior direction; their respective volumetric aspect ratio is 1:1.2:1.5, on average. Submandibulars display a similar aspect ratio (1:1.2:1.6), which confirms the slight elongation along the superior-inferior direction. The oral cavity is largest along the medial-lateral direction and smallest along the superior-inferior direction.

The parotid slab volumes of table 3.2 (i.e., median of 29.5 cm^3) are consistent with parotid volume estimates reported in the literature, which are generally within 25-35 cm^3, e.g., [74, 144, 145, 146, 147]. However, some reports differ, having either substantially higher (e.g., 43.1 cm^3 [148]) or substantially lower parotid volumes (e.g., 17.7 cm^3 [75]). In all studies described here, the range or uncertainties reported encompass the slab volume median of 29.5 cm^3. While this does not directly imply overall ROI consistency with the literature, it suggests both the lack of gross differences and some contouring practice congruence with other centres.

Many ROIs are missing from tables 3.2 and 3.3 (n.b. ROI counts are not equivalent). There are many reasons this can happen, including surgical removal (most common for neck dissections), poor tissue contrast, and treatment demarcation irrelevancy, such as when the primary tumour is within a salivary gland and sparing is impossible. Some patients are born missing one or more organs ('congenital aplasia' or 'agenesis') or with intact glands in abnormal locations ('aberrant' or 'ectopic') that are not detected during routine imaging [149, 150, 151].

3.5 Summary and Conclusions

BCCA head-and-neck cancer imaging and tissue demarcation practices appear to be typical. The parotid is indeed the largest salivary gland and is most widely contoured (∼72% of patients) followed by the submandibular (∼65%) and lastly the oral cavity which is not often contoured (∼29%). The routine CT imaging protocol has remained static for the decade from which the cohort considered in this work was derived.

Chapter 4

Salivary Gland Morphology and Topology

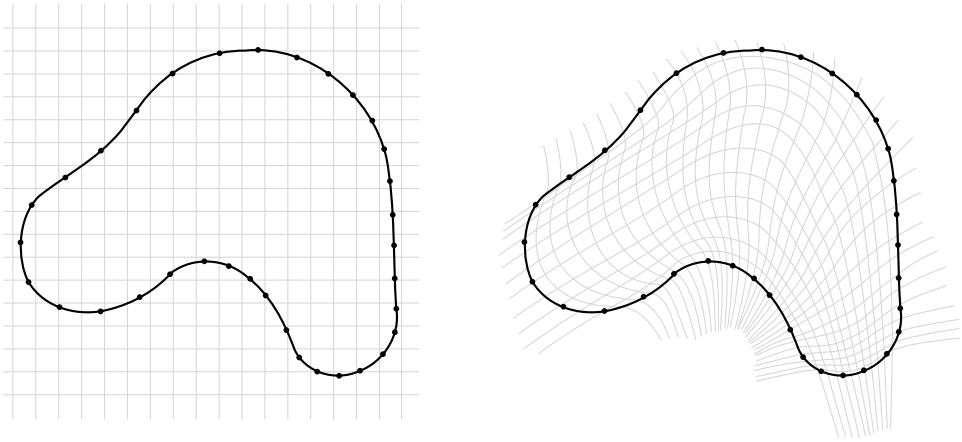


Figure 4.1: Examples of a non-conforming (left) and a conforming (right) coordinate system superimposed over a single parotid contour. The contour can be traced in the conforming (i.e., adaptive) coordinate system locally by translation along individual coordinate directions. The conforming system is a semi-conformal (i.e., angle-preserving) mapping of $\mathbb{R}^2$ Euclidean space.

A key requirement for performing inter-patient or inter-parotid analysis is defining a consistent spatial mapping between parotids. Parotids vary

considerably in size and shape. Not being able to consistently resolve parotid fine internal structure in CTs (i.e., ducts, vasculature, nerves, and lymph nodes) presents a major impediment for assessing regional effects. Development of a systematic mapping or correspondence is required not only for comparing like-regions within the parotid, but also for later understanding the association between critically important regions and anatomical structures or parenchyma.

Figure 4.1 shows grid lines for Euclidean space mapped into a single contour in two ways. While the non-conforming coordinate system (on the left) provides a valid mapping, it is not sufficient for inter-parotid analysis. The issue is that there is no notion of similarity when it is applied to different parotids. Intuitively, a suitable mapping would be consistent, bijective[1], and respect topological features, meaning a point along the mandible in one parotid should map to an equivalent point along the mandible in another parotid, and vice versa, for *any* two parotids. The conforming coordinate system on the right of fig. 4.1 meets these three criteria. However, this coordinate system is merely a sketch of what such a mapping might look like; development of a specific mapping procedure remains to be addressed. Furthermore, this illustrative mapping is for a single contour but a volumetric mapping is needed.

Parotids exhibit a distinct topology that is challenging for development of a consistent mapping. Two-manifolds (i.e., watertight surfaces) can always be triangulated, so assuming ROI contours are not directed (i.e., the orientation of individual contours is ignored, and all contours are homeomorphic to a disc, so that there cannot be annuli or holes in individual contours), it will always be possible to create a two-manifold regardless of ROI details [152]. Figure 4.2 shows orthographic projections of a single parotid from several viewpoints aiming at the characteristic 'saddle surface' which wraps around the mandible. The inferior 'tail' (toward the bottom of the page) and lateral

[1]Or, if not bijective, then at least surjective, meaning that the entire domain of the parotid must be covered, but there can portions of the domain of the map that do not map to the parotid. In other words, there can be disused elements in the domain of the map, but there cannot be any 'spares' in the domain of the parotid. Bijectivity would present a superior mapping as it would link the domains in a more meaningful way.

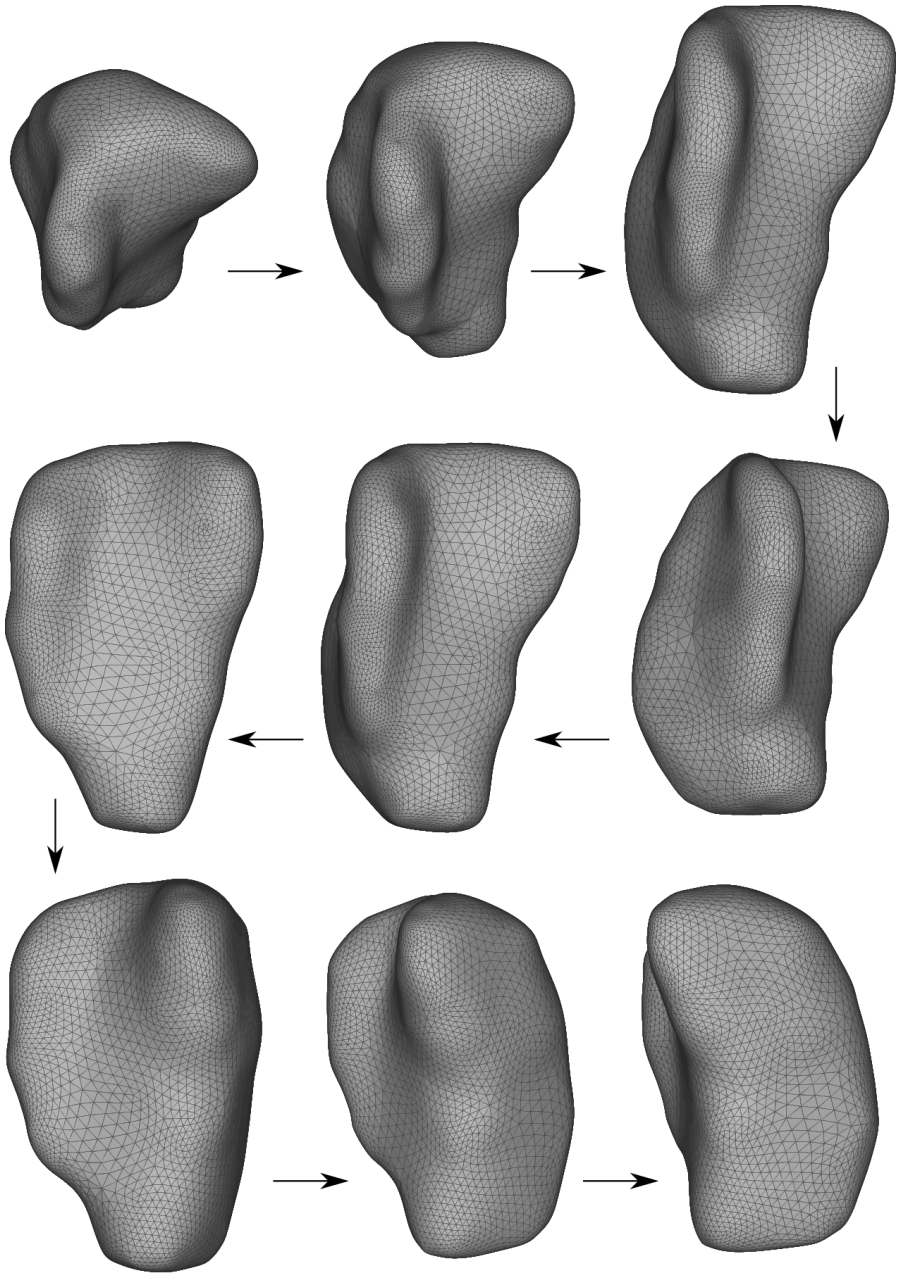


Figure 4.2: Topology of a typical parotid gland as clinically contoured at the BCCA from several viewpoints aiming at the 'saddle point.' Arrows trace sequential rotations of a single parotid gland. Top row: superior view panning to medial-anterior view; middle and bottom rows: medial-anterior view panning to lateral view.

and medial extrema (right and left of the page, respectively). Constructing two-manifolds from ROIs improves topological consistency, making it easier to identify characteristic features (n.b. compare fig. 4.2 with fig. 2.2). But also the triangulated two-manifold (i.e., homeomorphic to a hollow sphere) will bound a compact, oriented three-manifold (i.e., homeomorphic to a ball) which is homeomorphic to $\mathbb{R}^3$ Euclidean space [153, 154]. This implies that triangulated two-manifolds are in principle *capable* of supporting consistent, bijective mappings that respect topological features – all that remains is to actually find such a mapping using tessellated parotid two- or three-manifolds. As will be demonstrated in the following sections, this problem is difficult in practice.

4.1 Space-Filling Curves

Continuous space-filling curves that surjectively map every point in a confined region in $\mathbb{R}^2$ (usually a unit square) to a position on a line segment are well-known [155, 156]. (They are however surjective and not bijective because they are continuous and self-intersect.) Space-filling curves are special cases of fractal constructions, and though they are continuous they are not smooth, and are thus everywhere nondifferentiable. Variations exist[2] that fulfill dimensionality reduction, such as space-filling trees [157], rapidly-exploring random trees [158], and others [159]. More generally, they appear to be able to map arbitrary topological spaces (i.e., two- or three-manifolds) to a line segment (or at least be 'stitched together' to do so), though the literature is sparse on this topic, particularly for polyhedra in dimensions greater than two [160, 161, 162]. In particular, most polygon-filling curves rely on parametrization, optimization (such as simulated annealing), or both, which would make curve construction non-deterministic and therefore potentially[3] unsuitable for intra-organ comparison [163].

Space-filling curves can be constructed so that locality of position along

[2]Note that space-filling polyhedra (and the related concept of 'packing') are entirely unrelated from space-filling curves and trees.

[3]Wasser et al. [163] describe a heuristics-based method that may result in a sufficiently-deterministic construction.

the line segment reflects locality in the topological space (i.e., 'spatial clustering') which would be beneficial for intra-organ comparison. Hilbert [156] introduced such a curve more than a century ago, and it is thought to provide the best clustering of any known space-filling curves [164]. Construction iteratively becomes infinitesimally small, and iterations are known as (open) 'approximating polygons.' The first few approximating polygons are shown in fig. 4.3.

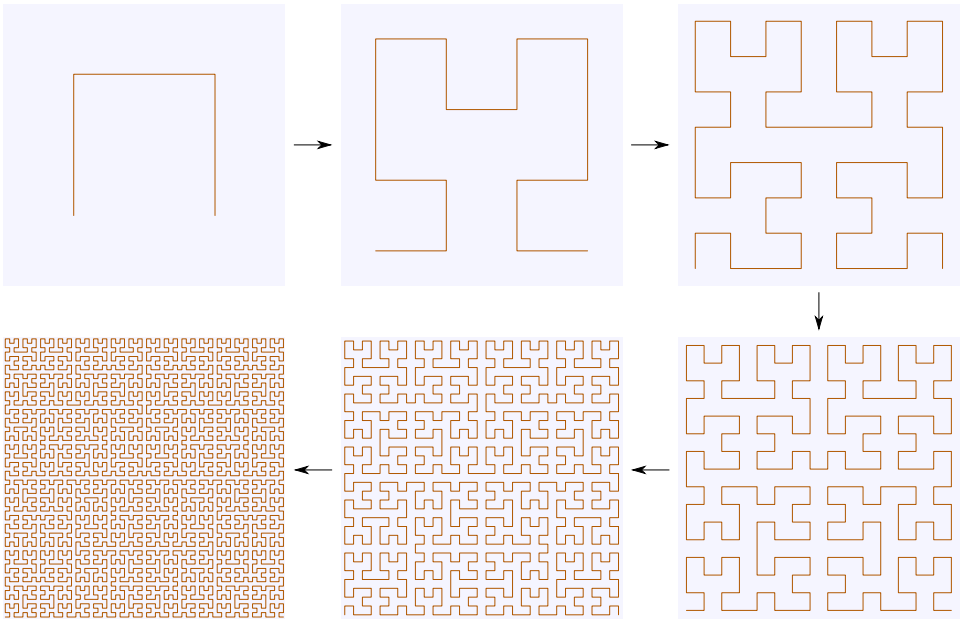


Figure 4.3: Construction of a space-filling Hilbert curve (i.e., open approximating polygons of increasing order). In the limiting case, every point in the unit square is surjectively mapped to a line segment. Locality is approximately preserved.

Spacing-filling curves have a variety of surprising uses, including applications in optimization [165], manufacturing [166], and scheduling [167]. In the present case, development of a locality-preserving curve that can map parotid three-manifolds was investigated. Despite being capable in principle of programmatically traversing parotid three-manifolds, they can not generally guarantee intra-parotid consistency, especially when parotid morphology differs substantially. It is not known if locality/spatial clustering

can be guaranteed when extended to polyhedra in dimensions greater than two. Furthermore, it is known that there are no continuous bijections from $\mathbb{R}^2$ to a line segment [168], so either bijectivity or continuity must be sacrificed – both of which would be convenient for inter-parotid assessment of intra-parotid regional effects. An effort was made to develop criteria to define suitable space-filling curves, but was ultimately abandoned in favour of other approaches. Helpful starting points for pursuing this approach are the framework for recursively generating multi-dimensional space-filling curves of Jin and Mellor-Crummey [169], the two-manifold density-based Hilbert curve resampling parameterization method described by Quinn et al. [160], and the heuristically-guided polygon-filling approach described by Wasser et al. [163].

4.2 Barycentric Coordinates

Barycentric coordinates can be used in simplicies to provide smooth mappings that conform when the hull is altered [170]. This feature, conceptually, makes them attractive for intra-parotid comparison. They are commonly used to interpolate within simplicies in which some quantity is known at the vertices but not within the simplex. They can also be used as a generic coordinate system within a simplex which abstracts the underlying coordinate system (e.g., Euclidean). Using this abstraction, they provide a means independent of the coordinate system of computing distance and can simplify integration (e.g., of partial volume, or radiation dose) [170]. The theory of generalized Barycentric coordinates which extend to n-sided polygons (i.e., in $\mathbb{R}^2$) is well-known [171], but extensions to arbitrary n-polytopes (i.e., in $\mathbb{R}^3$ and higher) appear to be limited only to convex polytopes [172, 173]. This means the method is only applicable in the present ($\mathbb{R}^3$) case to simplicies since salivary gland ROIs define, in general, non-convex shapes. Conceptually similar methods, such as Wachspress basis functions [174], similarly appear to be limited to convex polytopes [175]. So-called 'mean value coordinates' are able to interpolate (and even smoothly deform) arbitrary n-polytopes, but do not appear to be usable as a coordinate system due to lack of isomorphism [176, 177]. Even if they could be adapted to provide a coordinate system,

they do not appear to generalize easily when the intra-parotid correspondence between ROI vertices or 2-manifold facets are unknown (which is precisely the problem that these methods are supposed to address). Compared to other interpolating approaches, Barycentric coordinates over simplicies are therefore the most viable way to generate a vertex-based coordinate system.

In $\mathbb{R}^2$ simplicies are triangles and in $\mathbb{R}^3$ simplicies are tetrahedrons; tetrahedrons individually are woefully incapable of representing generic ROI-derived three-polytopes, which may comprise thousands of facets. On the other hand, if the equivalent three-manifold is tessellated with tetrahedrons, then points within each individual tetrahedron can be smoothly mapped, and all that is needed to achieve a consistent bijective intra-parotid mapping is to find a correspondence between intra-parotid tetrahedrons[4]. This correspondence problem presents many difficulties: tessellation can be made to produce dissimilar three-manifolds with differing numbers of tetrahedrons themselves with distinctly different positions and orientations. Any correspondence will need to be robust to all three differences, but still provide consistency and bijectivity. Deriving a suitable mapping by finding a correspondence without additional information is unlikely due to the complexity. Auxiliary information in the form of topological landmarks present a pathway out of the quagmire. If correspondence between tetrahedrons is eschewed for correspondence between landmarks, the problem is simplified considerably. If landmarks are suitably chosen, all that will be needed to establish correspondence are a list of volume-normalized geodesic distances[5] from each landmark,

[4]If this is confusing, consider that this technique shares many similarities with finite element methods (a.k.a. finite element analysis). The aim of finite element methods are to solve for a field (usually by solving a differential equation) over a tessellated three-manifold. However, finite element analysts may be confronted with the present problem if they have solved a demanding problem over one geometry and wanted to extend it to another similar geometry without having to re-solve the problem. In other domains the problem may be referred to as a 'correspondence' problem or maximal similarity problem. It is also conceptually similar to 'deformable registration' which is discussed shortly.

[5]Be aware that these distances can not be represented as a vector because they do not form a linear space; for example, their sum is meaningless. Likewise, an arbitrary number of landmarks can be specified which will result in a list with greater or fewer elements than the number of dimensions, leading to an under- or over-specified problem. If landmarks are carefully chosen (e.g., to avoid collinearity) then an exact location may be found. In general, though, an as-close-as-possible system should be employed to improve robustness,

e.g., x units from landmark A, y units from landmark B, and z units from landmark C. Then Barycentric coordinates can be employed to locate the corresponding point somewhere within a tetrahedron. Note that landmarks in this system could be anywhere in either the ROI two- or three-manifold.

As already noted parotids present several topological landmarks, such as the saddle surface, tail, and other extrema (see fig. 4.2). However, not only do parotids vary in volume and size, as described in section 3.4, but the overall *shapes* can be inconsistent, even in the same individual. It is therefore not reasonable specifically to specify features, not only because they may be absent in some parotids, but also because specification would naively require manual (and thus subjective) identification. This approach would not be consistent. However, an objective measure that *can* be used to identify topological features is *curvature*.

There are multiple types of curvature for two-manifolds. 'Mean curvature' intuitively describes the change in surface area relative to the change in volume when a surface is deformed. It is defined as the mean of the principal curvatures[6] – the largest and smallest curvature of all possible $\mathbb{R}^1$ slices that intersect a given point [178]. Another, Gaussian curvature, is the *product* of principal curvatures. Discrete approximations of curvature can be computed on two-manifolds using the method described by Meyer et al. [6]. It is known that saddle surfaces have negative Gaussian curvature [179], and that the mean curvature of symmetric saddles should be zero (which may be easier to algorithmically detect), but otherwise it is not *a priori* clear which curvature will best isolate the distinctive parotid features. As can be seen in fig. 4.4, mean curvature more uniformly highlights the saddle surface, ridges, and protruding lobes than Gaussian curvature. It is therefore used for visualization purposes. Ultimately, the most general approach is to create several derivative curvature maps tailored to select specific features.

which means that a true coordinate system and exact correspondence may not be possible. Global positioning systems face similar issues and provide good enough geolocation for most purposes. Likewise, for the present purposes a scheme employing geodesic distances may be good enough.

[6] Principal curvatures are more precisely defined as the inverse of the radius of the largest and smallest osculating circles at a given point on a two-manifold.

Indeed, measures other than curvature can also be used, such as the 'shape diameter function' which estimates the local diameter and can be used to locate protruding lobes and the tail [180].

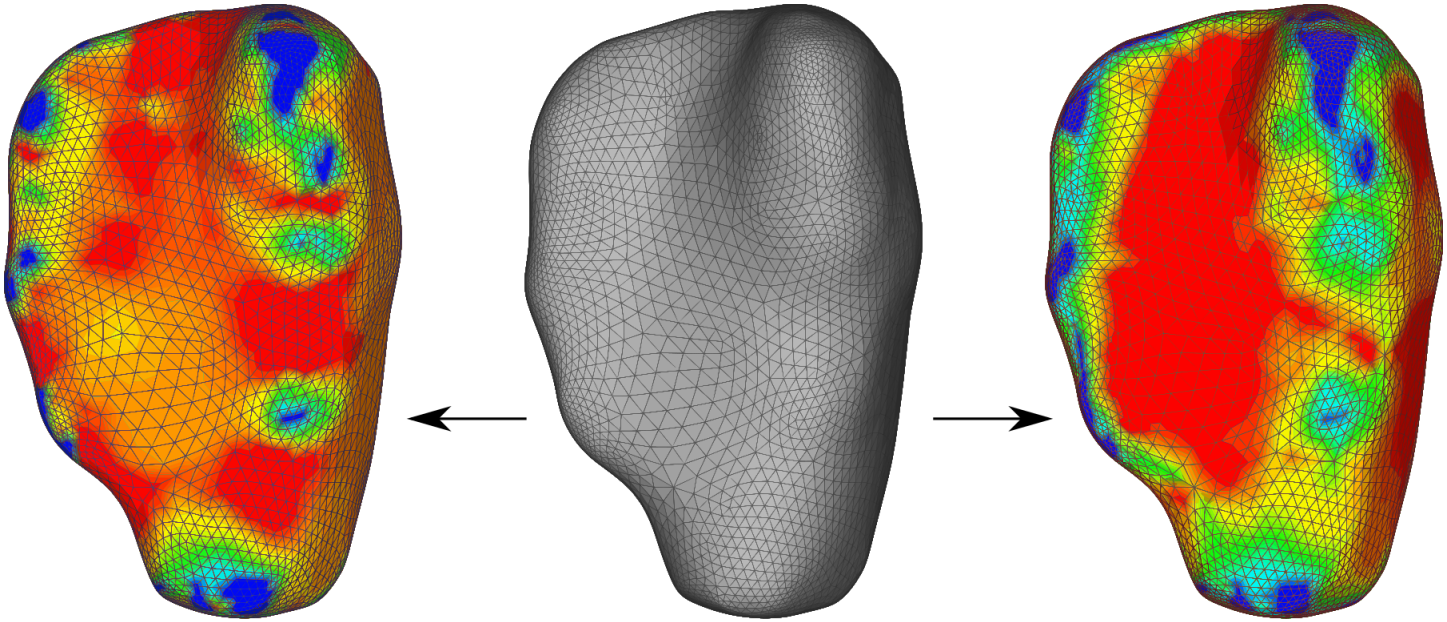


Figure 4.4: A parotid demonstrating two kinds of curvature (Gaussian on left, mean on right) which were computed using a discretization scheme described by Meyer et al. [6]. Red and blue represent mean curvature extremes (maximum and minimum). Mean curvature highlights the saddle surface more intuitively than Gaussian curvature, but neither consistently highlight the ridges.

Landmarks must be consistently and precisely located. If they are not, the internal coordinate system will be tainted by the imprecision. In other words, the method is not robust to landmark specification errors. Therefore, detection of landmarks from derivative curvature maps (and other maps) needs to be robust. Clustering techniques can be used to locate curvature-connected groupings over the two-manifold using geodesic distances. Then representative points, such as a curvature-weight averaged vertex position projected back onto the two-manifold, or the point within the cluster with maximal geodesic distance from the boundary, can then be taken as the landmark (e.g., the saddle point at the centre of the saddle surface).

Performing this procedure with the 'Density-based spatial clustering of applications with noise' (DBSCAN) clustering algorithm [181] and a variety of curvature derivative maps lead to landmarks with inconsistent positions. Mean curvature for six pairs of parotid glands is shown in fig. 4.5; it is clear that there is substantial variation in the position and extent of features. In particular, the saddle surface has a disperse spatial extent, which leads to localization errors when clustering, and ridges and protruding lobes do not present consistent curvature. In addition to normal intra-parotid variations, curvature (and all quantities derived from two-manifolds) were found to strongly depend on: (1) contouring minutiae, (2) the method of triangulation, and (3) any post-processing performed on the triangulated mesh, such as refinement/subdivision. (Figures 4.2, 4.4 and 4.5 were all triangulated using Delaunay triangulation and two iterations of Loop's subdivision method [182, 183].)

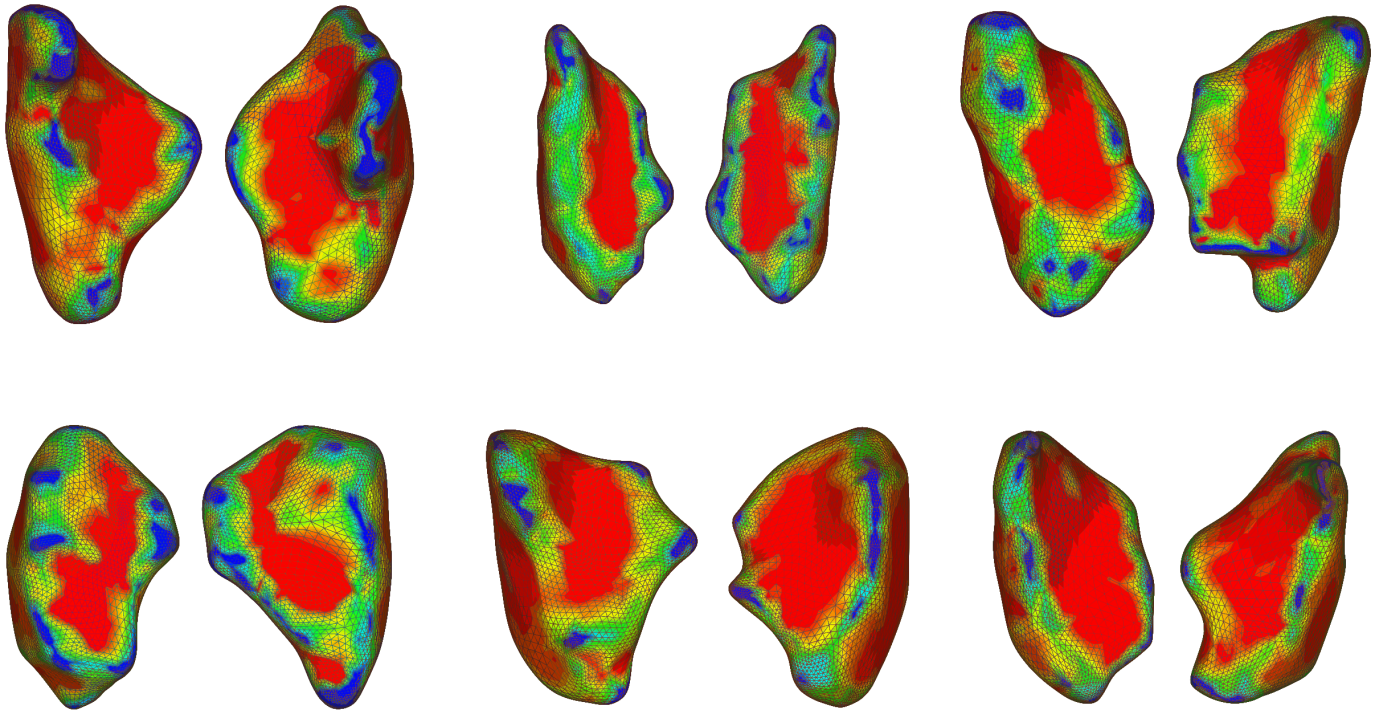


Figure 4.5: Mean curvature for six parotid gland pairs. Features are easy to visually identify, but difficult to robustly detect and locate computationally due to natural variations.

Furthermore, these problems can not easily be overcome; there is inherent ambiguity in two-manifold reconstruction using clinically-defined ROI contours. Figure 4.6 demonstrates two ambiguities that can strongly affect surface shape and curvature. To this end, the clinical Digital Imaging and Communication in Medicine (DICOM) standard provides a supplemental standard for storing two-manifolds directly, without conversion to planar contours [184, 185]. However, uptake by vendors has been slow and it is not used in the BCCA.

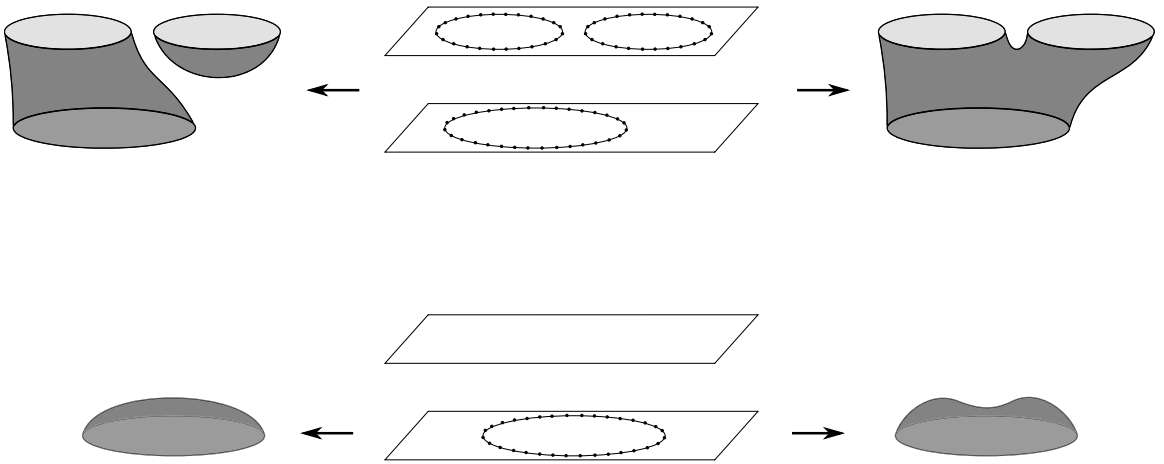


Figure 4.6: Demonstrations of topological ambiguity due to clinical tissue demarcation using ROIs. Top: ambiguous connectivity between image slices – both are valid and could be connected such that a 2-manifold homeomorphic to a sphere is produced. Bottom: ambiguous curvature for extrema contours – both satisfy the terminating boundary condition imposed by the adjacent slice.

The specific failings of two-manifold-based approaches are hard to ascribe to any single factor. It is likely that triangulation ambiguities can be overcome by consistently triangulating all parotids in the same way. However, the approach is inherently sensitive to the ROI *surface* rather than the *volume*. This means it is likely sensitive to factors that affect contouring, such as image contrast, window and level settings, variations in individual contouring practices and quirks, contouring subjectivity, and possibly even vertex density or the total number of vertices used in individual contours. Use of non-curvature measures may decrease reliance on surface minutiae. However, seemingly more robust volumetric measures defined over two-manifolds, such as the shape diameter function, may similarly be affected by contouring minutiae; as can be seen in fig. 4.5 the local diameter of protruding lobes is extremely variable and likely also subjectively contoured. Furthermore, a landmark-based approach is unable to handle parotids with missing landmarks. So a small number of landmarks, the 'lowest common denominator,' must be identified beforehand. This implies that the technique may fail to be applicable to ROIs added at a later date, which severely limits generalizability of the approach. It also reduces the robustness of landmark specification. Work on two-manifold methods including Barycentric coordinates was thus discontinued in favour of methods involving three-manifolds or that do not require tessellation.

4.3 Segmentation

A more robust method of defining position and locality within the parotid is to (1) temporarily do away with the notion of a coordinate system, and (2) recursively sub-divide the *volume* into discrete sub-volumes. If infinite recursion is permitted, then in principle a coordinate system based on infinitesimal bounding volumes could be reconstructed. However, infinite recursion, compared with finite recursion, would more strongly rely on contours because larger sub-volumes are more broad and therefore resilient to errors in translation. Recursively sub-dividing a small number of times will effectively

tile[7] the parotid. If sub-volumes are taken as atomic entities of space, then position and locality are discretized, but are also more robust to contouring discrepancies. For example, if recursive volumetric sub-division is performed then, *ceteris paribus*, as long as the volume remains the same, surface curvature has no bearing on sub-volumes that do not intersect the surface. The difference between volumetric and surface-based approaches is somewhat akin to the difference between the mean and median of a distribution – the mean is sensitive to every sample, but the median is only sensitive to rank.

Segmentation generically refers to the act of demarcating an object into smaller, more cohesive entitites. The term 'contour segmentation' was coined by the author to refer to the volumetric recursive sub-division tiling approach when it is applied directly to ROI contours without performing two-manifold tessellation[8] [7]. More generally, contour segmentation can be applied repeatedly – recursively or sequentially – to construct a tessellation of collections of solid shapes, including polygons embedded in $\mathbb{R}^3$. A tool for robust and composable segmentation of planar contours was constructed: `DICOMautomaton`. An ROI contour approach is taken, rather than two- or three-manifold approaches, which eliminates issues stemming from tessellation. Operating on planar contour polygons also enables 'lossless' reversible segmentation and is dramatically faster than surface reconstruction or bitmap/voxel-centric approaches [7]. No prior work was found in the literature, and `DICOMautomaton` continues to be the only tool capable of direct contour (sub-)segmentation (that the author is aware of).

Details on the specific segmentation methodology applicable to this thesis are provided in chapters 10 and 11, but a brief depiction of the various contour segmentation capabilities of `DICOMautomaton` are shown in figs. 4.7 to 4.10. Figure 4.7 demonstrates simple, single-contour tilings using radiation dose heuristics (left and centre) and a scale independent Cartesian tiling (right) reminiscent of the left side of fig. 4.1. Figure 4.8 demonstrates recursive segmentation, first using projective segmentation (n.b. uses ray-

[7] Note that in the following, 'tiling' should be taken to mean both proper and improper tiling, since vertices are not always shared by all adjacent sub-segments.

[8] Not to be confused with *image segmentation* or *mesh segmentation*.

casting – described in chapter 11) and then using per-sub-segment planar segmentation, resulting in a somewhat conforming tiling reminiscent of the nearly-conformal mapping of fig. 4.1. Figure 4.9 demonstrates more advanced recursive single-contour tilings, including (on the left) a semi-random, semi-periodic tessellation in which all internal edges are aligned in one of four randomly-chosen directions, (centre) a challenging optic chiasm tiling with internal edges aligned with Cartesian-axes, and (on the right) a telescoping segmentation of a left parotid contour with blocks of Cartesian grid arranged inside larger 'neighbourhood' sub-segments. Finally, fig. 4.10 demonstrates volumetric segmentation of whole left parotid ROI into: (bottom-left) a core and peel, (middle) medial and lateral volumetric halves via projective segmentation, and (top to top-right and bottom-right) planar segmentation into equal-volume sub-segments. In all cases contour area is used as a surrogate for volume (i.e., slab-volume).

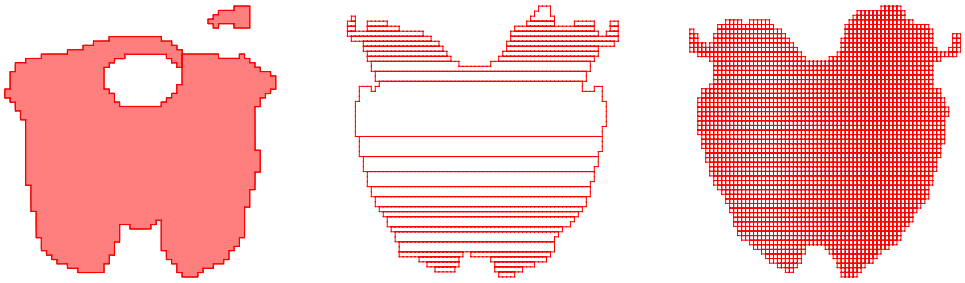


Figure 4.7: Tiling segmentations of a single contour of an axial 'body' contour at the shoulder level. From left to right: first, a heuristic segmentation based on absolute radiation dose, clearly demonstrating sparing of the spinal cord in the encircled region (n.b. adapted without modification from [7]); second, a heuristic segmentation based on spatially-varying heuristic based on the local dose gradient; and third, a recursive, scale independent Cartesian tiling.

4.4 Deformable Registration

An alternative to segmentation is deformable registration. Deformable registration methods are well-known, and can provide a way to map CT images,

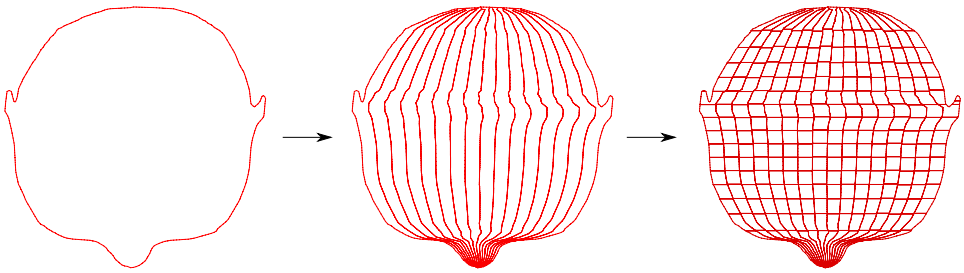


Figure 4.8: Recursive mixed segmentation that progressively tiles an axial body contour at nose level (left to right). The first segmentation is a medial-lateral projective segmentation (n.b. described in chapter 11) and the second is a per-sub-segment coronal planar segmentation. This figure was adapted from [7] and modified to simplify presentation.

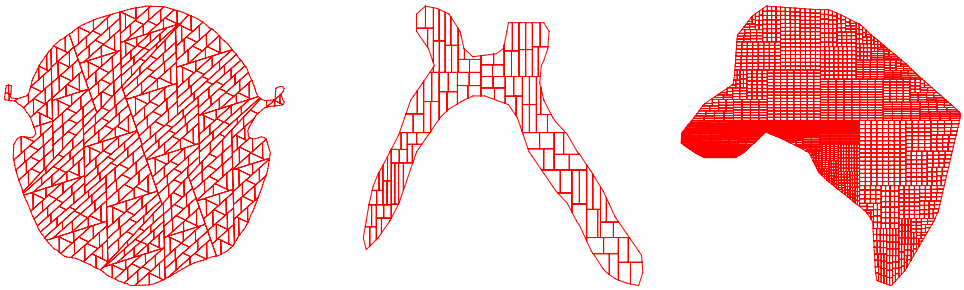


Figure 4.9: More advanced recursive single-contour tilings. From left to right: first, a semi-random, semi-periodic triangulation of an axial body contour at the level of the ear in which all internal edges are constrained to one of four randomly-chosen directions (n.b. adapted without modification from [7]); second, a challenging semi-Cartesian tiling tiling on a star-shaped optic chiasm contour; and third, a telescoping segmentation of a left parotid contour with blocks of Cartesian grid arranged inside larger 'neighbourhood' sub-segments. These examples highlight that recursion can be used to generate adaptive, arbitrary tilings within n-polygons.

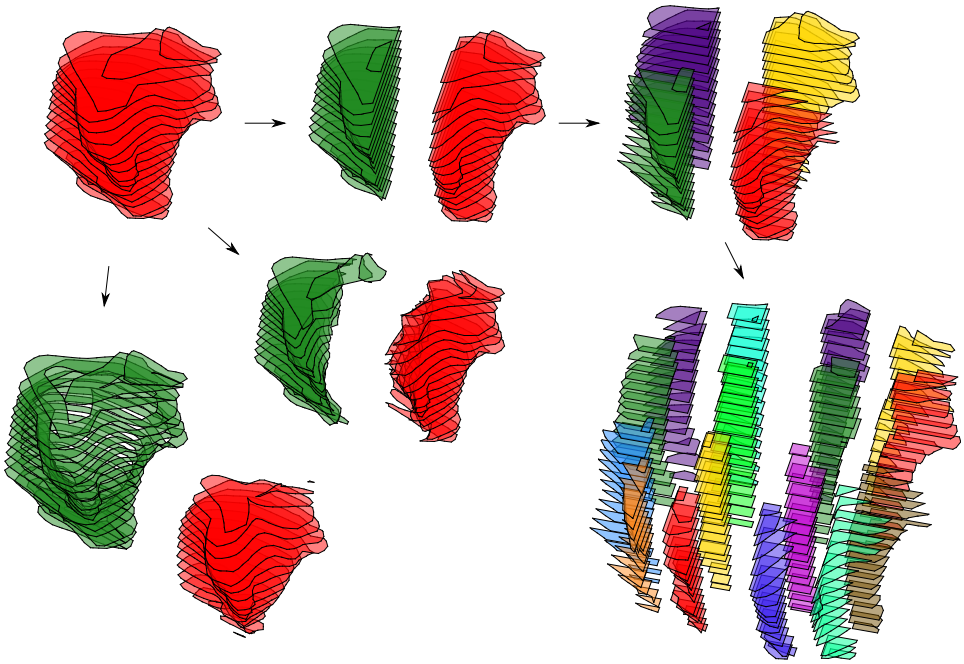


Figure 4.10: Volumetric segmentation of a whole left parotid ROI (top-left) into: (bottom-left) a core and peel, (middle) medial and lateral volumetric halves via projective segmentation (n.b. described in chapter 11), and (top to top-right and bottom-right) recursive or 'nested' planar segmentation into equal-volume sub-segments. These examples highlight that recursion can be used to generate adaptive, arbitrary tilings within oriented polyhedra which have been sliced to produce co-parallel planar contours. Parts of this figure were adapted from [7] and recoloured to simplify presentation.

dosimetric volumetric images, and ROI. There are several reliable registration algorithms that develop bijective mappings (i.e., 'diffeomorphic' algorithms; note that not all algorithms are diffeomorphic – see [186, 187, 188, 189] for discussion and [190] for a widely-available and popular diffeomorphic algorithm). Cross-registration can be used to register each parotid to each other parotid individually, but this will result in poor scaling with the number of ROI and will in general result in registration cycles that are not bijective (i.e., if a point in parotid A is mapped to parotid B, and then that point

is mapped to parotid C, and then that point is mapped back to parotid A, it will in general not coincide with the original point). These problems are overcome by deforming all parotids to a single 'prototypical' parotid and then performing all mappings using the prototype as an intermediary. However, while this approach solves the cycle non-bijectivity problem, it causes a specific dependency on the prototype. As seen in fig. 4.5, parotids are highly variable and selection of a representative prototype – no matter how appropriate – will inevitably present ramifications for the deformation quality of dissimilar parotids. Additionally, not all registration algorithms are 'symmetric.' Symmetric algorithms produce the same mapping whether image A is registered to image B or vice versa, but non-symmetric algorithms do not necessarily. Non-symmetric algorithms clearly spoil cross-registration registration cycle bijectivity. Segmentation, in contrast, is prescriptive, meaning that bijectivity is tautologically maintained in all cycles and no prototype is needed.

4.5 Conclusions

Space-filling curves, Barycentric coordinates, registration, and segmentation are all potentially viable techniques for investigating regional effects. Each have specific trade-offs. Space-filling curves are in some ways the least-constrained approach, but will require an unknown amount of work to develop curves that are applicable to ROI and describe parotid morphology and topology in a meaningful way. Barycentric coordinates suffer from sensitivity to contouring and ROI two-manifold tessellation ambiguities, and the approach ultimately relies on subjective assessment of landmarks. Deformable registration requires selection of a prototype to guarantee bijectivity when intra-parotid analysis is required, but is otherwise a flexible and general approach. Segmentation generally requires use of non-standard spatial localization and locality (e.g., using the adjacency of sub-segments), but otherwise presents a minimally-subjective, flexible paradigm that can be adapted to almost any segmentation problem. Given these findings, and owing to the versatility, simplicity, recursive generalizability, and avoidance of prototypes,

segmentation was selected for assessment of regional effects.

It is worth mentioning that a space-filling approach to segmentation – tree-based methods, in particular – would provide the most flexible approach. This route was not pursued because segmentation based on volumetric specifications was found to be sufficiently flexible on its own.

Chapter 5

Why are Salivary Glands Irradiated?

5.1 Primary Cancers Within Salivary Glands

Cancers of the salivary glands themselves are uncommon, accounting for ∼6% of head-and-neck cancer incidence and ∼0.3% of all cancer incidence [191]. Benign tumours are also rare [192]. When cancers do occur, surgery followed by postoperative radiotherapy is common and leads to a reasonable amount of local tumour control [193].

It is self-evident that patients undergoing radiotherapy for salivary gland cancers would receive substantial salivary gland irradiation. However, *almost all* head-and-neck cancer patients treated with therapy receive non-negligible salivary gland dose. There are two key reasons. First, head-and-neck anatomy is complex, and given the distribution of salivary glands, some irradiation is inevitable. Second, healthy tissues (both surrounding and distant to the primary disease site) are often irradiated to reduce risk of additional primary cancers. These two topics and some related issues are the focus of the remainder of this chapter.

5.2 Practical Head-and-Neck Anatomical Constraints

As discussed in chapter 2 minor salivary glands are distributed throughout the oral cavity and upper aerodigestive tract. Major salivary glands wrap around, behind, and below the mandible, shielding much of the nasopharynx and oropharynx from lateral radiation. The brainstem and spinal cord awkwardly block the same anatomy from posterior radiation, and radiosensitive visual structures block superior-anterior radiation. Superior access for radiation delivery is, in almost all head-and-neck cancers, blocked by the brain[1] (which is bulky and sensitive to radiation) or brainstem (a crucial structure that can cause severe complications when irradiated). Even if the presence of these organs is ignored, radiation delivered from a superior position along the length of the body would result in a deeply-penetrating dose profile. Additionally, recurrence in the primary tumour site is fairly common at ∼12% and often requires aggressive treatment with additional spatial margins to eliminate microscopic spread of disease [194]. Therefore, salivary glands are irradiated as a geometrical consequence of the complex head-and-neck anatomy in order to spare more critical tissues and obtain local control over the primary disease site.

5.3 Metastases and Second Primary Cancers

The majority of head-and-neck cancers are squamous cell carcinomas [195]. Initially, squamous cell carcinomas are confined from invading adjacent tissues by basement membranes or other fascial tissues (a.k.a. "squamous cell carcinomas *in situ*") [196]. In some cases these encapsulations remain intact, but most often when the disease penetrates these encapsulations (i.e., it becomes 'invasive') it is able to metastasize, spreading throughout the body creating 'second primary cancers' [197]. Metastatic tumourigenesis proceeds from progenitor cells originating within the primary tumour site [198]. Most cells cannot withstand incompatible microenvironments in other anatomical regions, but distant tissues may begin to support cells with compatible

[1] Note: brain cancers are not classified within the head-and-neck cancer category.

phenotypic abnormalities [199]. Tumours are thus spread, but biological characteristics of the metastasizing cells are retained. However, they can become compatible by gradual diversification through tumour evolution or progression, and thus tumours can adapt to their environment [198].

Survivors of squamous cell carcinoma head-and-neck cancers face life-long risk of new primary cancers in addition to frequent aerodigestive tract and cardiac illnesses [200]. The risk of developing additional primary tumours is 3-5% per year, and all sites in the aerodigestive tract are likely [201, 202]. Five-year survival varies substantially (20-90%) relative to those without earlier cancers depending on the origin site and progression [13], but overall five-year survival is around 40-60% for combined head-and-neck cancers [195, 203, 204, 205].

Many patients with squamous cell carcinoma head-and-neck cancers initially present at a relatively advanced stage due to ongoing difficulties with early detection [206, 207]. Given the grim survival prospects once a tumour has metastasized, generally high recurrence rates, issues with detection methods, advanced disease progression before diagnosis, and chronic difficulties in treatment due to critical structure placement, preventative measures are often taken to reduce the risk of second primary cancers. the easiest measure is to aggressively irradiate cancerous tissues, but this impinges upon salivary gland toxicities.

5.4 Lymph Nodes Must be Irradiated

Cancerous cells can spread to lymph nodes near the primary tumor (referred to as 'nodal involvement') which will distribute them throughout the body. Some will eventually metastasize; carcinomas most often metastasize via this lymphatic pathway [208].

Lymph nodes are distributed throughout the head-and-neck ('cervical lymphatics') [209]. Some are proximally inferior to the parotid gland; others are proximal to the submandibular gland. Submandibular glands are removed along with lymph nodes during (standard) radical neck dissections due to frequent lymph involvement [209]. Because lymph nodes need to be irradiated

to destroy cancerous cells, parotids often receive radiation dose regardless of the primary tumour site. Patients without metastases are treated anyways, unnecessarily, because of difficulties assessing metastatic state and high risks associated with loss of local control [210].

Cervical lymph nodes (throughout the neck) are typically wholly resected during neck dissections [210]. However, even in this case, or if proximity to major lymph nodes in the neck is ignored, there is frequent lymph node involvement in the parotid. In some cases node infiltration is a common condition in its own right (e.g., Sjörgens syndrome) and it is therefore sensible to irradiate the parotid to protect against lymphatic infiltration and metastatic transport. Thus, the proximity of lymph nodes often necessitates some salivary gland irradiation, even when the tumour site is distant.

5.5 Head and Neck Cancer Epidemiology

Lifetime risk of cancer for American citizens in any site is ∼40%, and mortality rates due to cancers surpass heart disease in those younger than 85 years of age [211, 212]. In 2005, globally the head and neck region was the sixth most common site for cancers. Head-and-neck cancers alone represented ∼6% of new cancer burden (∼650k incidences) and caused ∼350k deaths [213, 214].

There is considerable incidence variation around the world [13, 215]. This is thought to be due mostly to systematic and habitual exposure to carcinogens, especially smoking (or other tobacco products) and alcohol consumption, but several cancers present geographical, genetic, diet, and viral predispositions [13, 216, 217, 218, 219, 220]. Tobacco and alcohol have a 'knock-on' multiplier effect, and are implicated in approximately 3/4 of all squamous cell carcinomae [219, 221, 222].

Global head-and-neck cancer incidences have fallen slightly in recent years [195], but have risen in many populations, including Canada [13]. Recently, oral cancers kill one person worldwide each hour [223]. This toll is greater than cancers of the liver, kidney, brain, and gonads. Complications surrounding treatment of head-and-neck cancers, therefore, are frequently encountered in the clinic and remain a significant issue for newly diagnosed patients and

long-time survivors alike.

5.6 Summary and Conclusions

In almost all head-and-neck cancers the major salivary glands are irradiated to destroy primary cancer microscopic disease or to reduce risk of metastatic lymphatic transport. Modern technologies like Intensity-Modulated Radiotherapy Treatment (IMRT) and Volumetric Arc Therapy (VMAT), in which the radiation beam cross-section is 'sculpted' by moving jaws, or proton therapy are capable (or potentially capable) of delivering dose distributions that are tailored to selectively irradiate small sub-organ volumes. Therefore a more complete understanding of local resilience to radiation-induced salivary dysfunction will improve treatment planning risk assessment. Depending on the distribution of critical regions, this knowledge may permit irradiation of lymphatic structures and the primary disease site in such a way that the treatment efficacy is maintained but risk of toxicity is lowered.

Chapter 6

Salivary Dysfunction and Xerostomia

6.1 Introduction

Xerostomia is not itself a disease, but rather is a symptom of various medical conditions which presents as subjective dry mouth. It often results from diseases or trauma that cause salivary dysfunction, such as radiotherapy; medication (including common chemotherapuetic agents); and autoimmune diseases such as sarcoidosis (an inflammatory disease in which granulomas develop throughout the body), Sjögren's syndrome (a disease that suppresses function in mucous- and moisture-secreting organs), and rheumatoid arthritis [224, 225].

The impact of salivary dysfunction in cancer patients is multifarious. Advanced cases are known to severely reduce a patient's perceived Quality-of-Life (QoL) [226]. Even moderate, temporary dysfunction can impact a patient's primary faculties, reducing eating, sleeping, and communication to difficult, tedious, and painful exercises [148, 227].

As described in chapter 5, salivary glands are frequently irradiated out of necessity due to complex head-and-neck geometry and the confined presence of multiple organs at risk (e.g., brain stem, spinal cord, larynx and pharynx, oral cavity, and parotids). High doses of radiation are capable of completely

and permanently ceasing salivary function [10, 228]. However, the intensity of glandular damage can be curbed by controlling exposure, such as limiting the dose received and limiting the volume of irradiated tissues [40, 229, 230, 231, 232].

Radiotherapy dose profiles will sometimes irradiate a large volume and number of distinct tissues, including a variety of salivary glands distributed throughout the oral cavity. It is not *a prior* clear whether damage to large, concentrated glands or small, distributed glands will produce better patient outcomes. At a basic level, surgical removal of the submandibular has shown to increase risk of xerostomia [233]. Conversely, surgical transplantation of submandibular glands away from the target volume prior to radiotherapy can reduce xerostomia risk [234, 235]. Furthermore, mean dose to the accessory glands is a significant predictor of xerostomia, though not the most important [40]. This evidence suggests that xerostomia is not strictly an affliction of the largest glands, and that both submandibular and accessory glands contribute.

Nevertheless, the parotid gland is known to contribute the largest portion of stimulated whole-mouth saliva. The most heavily impacted primary faculties (i.e., swallowing, eating, speaking) are predominantly impacted by dysfunctional stimulated salivary flow[1] [227]. Additionally, the parotid almost always receives a significant dose in head-and-neck cancers while lesser glands, which by merit of their distribution, may be collectively spared. The parotid is therefore decidedly the most important gland and thus forms the basis for current clinical guidelines involving salivary glands [236].

In the remainder of this section, aspects of xerostomia and dysfunction are elaborated upon with an eye toward reducing their many facets to a level manageable for analysis of regional effects. Questions that remain open (or disputed) in the literature are provided heuristic answers that will allow analyses to proceed under the assumptions made. In particular, the following is addressed: the link between dysfunction and xerostomia, effective prioritization of dysfunction and xerostomia measurements (when both are available for analysis), whether short-term or long-term xerostomia is most relevant, and how to make headway when a surplus of dosimetric data is

[1] This is shown to be true in the BCCA cohort in section 6.4.

available.

6.2 History

Xerostomia – before it was known as xerostomia – was first considered a medical condition in its own right by Bartlet in 1868 [237, 238]. It was precisely defined and named from the Greek 'xeros' (meaning *dry*) and 'stoma' (*mouth*) in 1886 by Hutchinson [239]. He described severe xerostomia.

> *The tongue is red, devoid of epithelium, cracked, and absolutely dry, its appearance being like raw beef. The inside of the cheeks, the hard and soft palates, are also dry; the mucous membrane smooth, shiny, and pale. The salivary glands appear normal, and no mechanical obstruction has been detected in their ducts.* [...] *Articulation is difficult in consequence of the absence of moisture, and swallowing has to be assisted by constant sipping.* [...] *The disease reaches its greatest intensity suddenly, and then remains without change for years.*

Of particular note is that he drew a distinction between xerostomia (and the ramifications of having a dry mouth) and dysfunction. Indeed, salivary glands may produce or be able to produce saliva, but the saliva is either chronically inhibited or of insufficient quality (e.g., lacking an mucin content) to accomplish the normal functions of lubrication and protection.

Radiation-induced xerostomia is likely to be induced by different mechanisms than xerostomia induced by medication or immune diseases. The earliest recording of radiation-induced xerostomia was described in 1938 by Martin [240]. It may have been known earlier, though, as the report was accompanied with descriptions of secondary complications and remedies. Remediation of these complications has proven difficult and many remain problematic in modern times.

6.3 Complications

Xerostomia is a pathway for a diverse and numerous set of adverse effects. Besides the direct impact on primary eating, sleeping, swallowing, and speaking faculties, xerostomia is known to have a commanding impact on emotional welfare [241]. Even overall QoL is tainted by chronic pain resulting from xerostomia [226]. Dysfunction will often lead to increased opportunistic infection of the oral cavity which can itself cause severe complications. Specific complications that have been noted to result from xerostomia include: loss of taste, osteoradionecrosis, trismus (spasm of the jaw muscles causing the jaw to remain tightly shut), mucositis, and dental caries [242].

Dental caries are particularly prevalent and aggressive after head-and-neck irradiation [243]. This was well-known nearly 80 years ago; Martin [240] described the situation.

> *A complication occasionally associated with radiation xerostomia is a peculiar form of dental caries.* [...] *Beginning about two or three months after irradiation of the pharynx or of the oral cavity, metal fillings and inlays tend to loosen and fall out. In the cases of greater involvement the teeth may lose their natural glistening appearance and assume a dull, chalky hue. The substance becomes rather brittle and may wear away at the occlusal surfaces. Numerous cavities develop, especially near the gingival margin, so that the teeth tend to crumble or break off, leaving the root exposed at the gum level. Toothache is a prominent symptom. The direct cause of these dental complications is somewhat obscure.*

The modern belief is that radiation caries are caused by hyposalivation, though it has been suggested that diffuse radiation on the bones of the jaws and a reduction of blood supply through apical arteries could be to blame [244, 245, 246]. It is also thought that radiotherapy may exacerbate demineralization, which supports the loosening of fillings and bone deadening observed by Martin [240] [243]. Due to the potentially severe impact on general health, early recommendations included complete extraction of teeth prior to irradiation. This was generally to the detriment of the patient as

osteoradionecrosis and osteomyelitis of the mandible often followed extraction [247]. Nearly 80 years later, this practice remains in some parts of the world and the criteria for preventative dental extraction are still somewhat controversial [248, 249, 250, 251, 252]. Part of the issue is that post-radiotherapy dental extraction can directly lead to osteoradionecrosis, so there may be a benefit to pre-radiotherapy extraction in some cases [253].

Overall, it is *generally* accepted that dental extraction is warranted in patients with questionable prognosis or motivation [242, 254, 255]. If teeth are not extracted, modern preventative measures require heightened oral hygiene and include frequent application of fluoride solutions, limitation of cariogenic, sugar-rich or acidic foods, and application of artificial salivary agents [256]. There are also changes in dental practice, such as use of glass ionomer cements that have more favourable adhesive properties for radiation-stricken enamel and dentine [243]. Though a patient may be able to avoid such complication through diligence, effectiveness is limited by the patient's tolerance and rigor [242]. It is well-known that the majority do not follow pre-radiotherapy dental guidelines (see, e.g., [257]). Adhering to more stringent guidelines therefore seems unlikely for patients inflicted with the most severe or indefinite xerostomia, or who have been substantially and negatively impacted by complications. It is therefore difficult to develop a general course of action because the prognosis is multi-faceted.

6.4 The Association Between Xerostomia and Dysfunction

Patient-reported xerostomia, by definition, is subjective. Conversely, salivary flow can be measured quantitatively. It is clear that these two quantities are not exact surrogates. For example, xerostomia is a broad concept that encapsulates overall perception of the adequacy of salivary flow in distinct situations. It is also impacted by changes in salivary composition (e.g., viscosity) and the state of a patient's oral cavity mucosal lining (e.g., sore or infected) [225, 258, 259]. Conversely, salivary function can be impacted by factors that may not affect sensation of xerostomia, such as circadian

rhythms and hydration [25, 260, 261].

It should be unsurprising, then, that correlation between radiation-induced xerostomia and dysfunction has generally been found to be weak [10, 15, 40, 226, 262, 263, 264, 265, 266]. Differences are likely to stem from the aforementioned factors, difficulties in reliably measuring flow rates, and general perceptions of oral dryness [40, 263]. Conversely, some studies have found significant or mixed correlations [26, 226, 229, 267, 268]. It is unclear what the nature of the discrepancy is, but it could stem from use of multiple, inconsistent assessment instruments and techniques (n.b. discussed in sections 7.1 and 7.2). Unfortunately, despite weak-at-best evidence, salivary flow and xerostomia are frequently used interchangeably in the literature (e.g., [269]). Confusion is perpetuated by the use of various grading systems that intentionally conflate the two conditions for clinical purposes (see section 6.5). Besides surrogacy of the conditions, the implicit assumption of clinical grading schemes is that both conditions are 'switched on' by radiation, when in fact it is possible for a patient to develop xerostomia without dysfunction and vice versa [263]. Confusion of *radiotherapy-induced* xerostomia and dysfunction is less grievous than the general case owing to a correlation (however weak), and since salivary measurements likely are the most defensible objective surrogate for a subjective condition. But the practice is still technically conflation of two unrelated conditions and care should be taken to appropriately differentiate them.

6.4.1 Association in the BCCA Cohort

The correlation between post-radiotherapy objective flow measurements (W_b, W_{3m}, W_{1y}, and W_{2y} – whole-saliva measurements at baseline and three-months, one-year, and two-years post-radiotherapy, respectively) and subjective patient-reported QoL was assessed for the entire head-and-neck cancer patient cohort amassed at the BCCA over the decade spanning 2006-2016, omitting patients lacking data needed for each computation. The collection instruments and protocols are described in chapter 7, but in brief saliva is passively collected over a span of five minutes, and a nine-question question-

naire is administered (n.b. questions are described in section 7.1). Saliva was baseline-normalized by dividing baseline (i.e., pre-radiotherapy) saliva measurements so that patients receiving no radiation (and thus experiencing no induced dysfunction) would score 1.0 and patients with total loss of function would score 0.0. Individual QoL responses are of interval type (i.e., ordinal variables with equally-spaced divisions) and were transformed to the same range and 'orientation' by subtracting and dividing the maximum value (10 in all cases). Only saliva measurements and QoL data collected at the same appointment were considered (i.e., 'per-questionnaire' analysis, except for baseline normalization). Results for stimulated saliva are shown in table 6.1. Results for unstimulated saliva are shown in table 6.2. Patient attendance at follow-up appointments declines over time, so the number of datum compared are stated in each case.

Question Number	W_{3m}/W_b		W_{1y}/W_b		W_{2y}/W_b	
	r	N	r	N	r	N
2	0.232	524	0.063	296	0.198	182
3	0.303	522	0.171	298	0.178	183
4	0.310	525	0.150	306	0.210	183
5	0.325	521	0.221	306	0.283	183
6	0.205	528	0.155	307	0.267	181
7	0.328	525	0.188	305	0.234	183
8	0.320	529	0.149	306	0.225	183
9	0.213	527	0.156	308	0.123	183

Table 6.1: Pearson's correlation coefficients (r) between baseline-normalized whole-mouth **stimulated** saliva measurements and normalized and inverted individual QoL responses. W represents whole-mouth saliva, N is the number of questionnaires available. QoL instrument questions are described in section 7.1.

Both table 6.1 and table 6.2 display overall weak correlation[2] between

[2]The r partitioning method recommended by Evans [270] is used, in which $r \in [0.20, 0.40)$ is 'weak' correlation. This approach was taken in lieu of explicit significance testing to avoid testing difficulties when N is large and conflation of (domain-specific) relevance with significance. See Taylor [271] for discussion of the pitfalls of testing r.

Question Number	W_{3m}/W_b		W_{1y}/W_b		W_{2y}/W_b	
	r	N	r	N	r	N
2	0.187	523	0.197	295	0.163	185
3	0.276	523	0.248	297	0.038	186
4	0.269	526	0.224	305	0.023	186
5	0.279	522	0.232	305	0.128	186
6	0.270	527	0.158	306	0.131	184
7	0.293	526	0.289	304	0.144	186
8	0.283	529	0.228	305	0.177	186
9	0.210	526	0.120	307	0.131	186

Table 6.2: Pearson's correlation coefficients (r) between baseline-normalized whole-mouth **unstimulated** saliva measurements and normalized and inverted individual QoL responses. W represents whole-mouth saliva, N is the number of questionnaires available. QoL instrument questions are described in section 7.1.

subjective QoL and objective salivary measurements in the BCCA dataset. Stimulated saliva correlate most strongly; the average correlation for stimulated saliva is 0.217 vs. 0.196 for unstimulated saliva. Early correlation is greater than later correlation; the average correlation for W_{3m}/W_b (i.e., baseline-normalized whole-mouth saliva during the three-month follow-up) was 1.30-2.21$\times$ that of W_{1y}/W_b or W_{2y}/W_b (stimulated and unstimulated).

Note that objective and subjective measurements have been normalized differently. Saliva measurements are rarely exactly 0g/5min for pre-radiotherapy patients, and there is no practical upper limit to the amount of saliva produced. It therefore is sensible to divide the baseline to give a patient-specific normalization that nominally ranges over $[0, 1]$. In contrast, subjective questions present an interval scale which naturally has both lower and upper limits that all patients are uniformly subjected to. Furthermore, while baseline flow is almost always non-zero, the median baseline questionnaire responses for questions 2-9 was 0 ($N = 1531$). The 75^{th} percentiles for each question were all $\leq$2 (out of 10) and the 90^{th} percentiles were all $\leq$5. By and large, baseline questionnaire responses are low and most frequently exactly zero,

and therefore cannot be used for *multiplicative* normalization. Recreating tables 6.1 and 6.2 by *subtractive* baseline normalization resulted in weaker correlations across both stimulated and unstimulated saliva, on average (mean $r = 0.206$ with no QoL normalization vs. 0.192 with QoL normalization). The 20^{th} and 80^{th} percentiles showed similar shifts. Comparison of the differences of r for individual questions showed that they increased by an average of 0.015 ± 0.007 (mean $\pm$ std. dev. of the mean). Differences were symmetrically distributed about the mean. Overall, *subtractive* normalization is thought to be inferior to foregoing normalization altogether. The topic of normalization and how it impacts analyses is further elaborated upon in section 6.7.

Table 6.1 and table 6.2 also demonstrate that missing datum are almost always due to patients increasingly foregoing follow-up appointments as their treatment grows more distant in time. Xerostomia/dysfunction assessment is performed as part of a more general dental examination, and it is rare for patients to attend a dental examination and decline xerostomia/dysfunction assessment, even when they can not produce an appreciable amount of saliva over five minutes of stimulation. Rather, patients who forego follow-up appointments most often cite logistical reasons or being unable to attend due to work or family commitments [272]. Therefore, it seems likely that there are no *major* systematic biases linking severity of toxicity and assessment instrument obstruction or patient-censored data. Further exploration of this topic is presented in section 8.3.

6.5 Grading

Precisely what constitutes 'severe' dysfunction is patient-dependent. For example, a reduction in salivary flow by 5g/5min may be significant for one patient but not another. In cases where it is rapidly induced by an external factor, baseline normalization is commonly performed as in section 6.4. Grading – the process of converting continuous salivary flow measurements into a small number of interval quantities (and 'stages' or 'grades') to aid clinical judgment – is commonly applied using the analytic component of the Late Effects Normal Tissue - Subjective, Objective, Management, Ana-

lytic (LENT-SOMA) scale.

In the LENT-SOMA system, induced xerostomia and salivary dysfunction are conflated. Severity is divided into four grades, the worst of which represents reduction of $\geq$75% of the whole-mouth baseline salivary output. This is referred to as "severe" or "grade IV" xerostomia. As it is of primary use in the clinic and for developing clinical guidelines, much of the research literature do not precisely employ this scale. Sometimes the discrepancy is subtle. For example, whole-mouth saliva is occasionally replaced by single-organ (or single-organ-pair) output [273]. In this case "grade IV xerostomia" does not fully correspond to LENT-SOMA scale grade IV xerostomia because the subjective component is not single-organ-separable.

Several alternative grading systems exist, many with different origins. Many are reasonably widely used. Some are periodically updated and so multiple versions with small deviations can be found within the literature. Examples include the Common Terminology Criteria for Adverse Effects (CTCAE) system (versions: 1.0 from 1983, known as simply the 'CTC' system; 2.0 from 1998 [274]; 3.0 from 2003 [275]; and most recently 4.0 from 2009 [276]) which began as a means of grading adverse effects of chemotherapy [277], a system employed by the Danish Head and Neck Cancer Group (DAHANCA) [266], Dische's early system [278, 279], and the Radiation Therapy Oncology Group (RTOG) EORTC [280] late morbidity system. Studies using outdated versions can be found within the literature, which can muddy interpretation and direct comparison (e.g., a study from 2013 using CTCAE version 2, from 1998, which was two versions behind[3] [281]). Translation between systems is not always obvious or, in some cases, possible [282]. For example, the CTCAE scale represents grade 3 dry mouth in terms of *absolute* flow, rather than *relative* flow (see e.g., [283]) and may therefore be more suited for naturally occurring or slowly-induced dysfunction. Jensen et al. [266] provides a good but slightly dated guide to many such systems. The LENT-SOMA scale should generally be preferred as reports indicate it is

[3]Note that there are many reasonable ways in which this situation can arise. The most obvious is intentionally using old versions to maintain consistency with earlier or long-running studies.

at least as good as alternative systems for scoring late radiotherapy-induced salivary toxicities [279, 282, 284, 285]. Note that like the LENT-SOMA system, most alternatives conflate dysfunction and xerostomia by treating them as a common toxicity.

Grading systems are not *directly* used in this work. However, they are important because they inform clinical recommendations and make interpretation by medical staff easier. Treatment planning guidelines are developed specifically to avoid the most severe toxicity risk, and the patient cohort considered in the present work were all subject to recommendations using such guidelines (i.e., [285]). It is therefore important to understand how population-level parotid radiation doses are impacted. For example, the distribution of mean parotid doses around clinical guideline thresholds will not be Gaussian because doses near to clinical sparing recommendations are more likely to be aggressively optimized to achieve the threshold. In analytic scenarios, grading systems can also be used to convert continuous saliva flow measurements to ordinal/interval or nominal variables, and subjective ordinal/interval reportings to nominal. These 'munged' variables can be used for statistical classification, to provide estimates of uncertainty (e.g., binning), or simply to reduce noise. In the present work, however, a sufficiently large cohort was available so regression was performed directly on continuous objective data.

6.6 Dysfunction and Recovery

Salivary function loss can be induced in many ways. Surgical resection of sialolith (calcifications) occasionally results in inflamed ducts. Atrophy may ensue, leading rapidly to dysfunction [16]. Sudden onset of ductal atrophy appears common in salivary glands. Dietary changes are commonly reported to induce some atrophy via necrosis of acinar cells – even over the span of a single day in rats [32, 286, 287]. The mechanism is thought to be related to changes in masticatory activity [287]. Similarly in humans, extended periods of masticatory inactivity resulting from autonomic denervation, intubation, ductal ligation, or trauma-induced fistulae (i.e., internal or external expression

of the duct contrary to the normal papilla) are known to induce salivary gland atrophy [288]. Glandular regeneration occurs after intubation-, ligation-, and fistulae-induced atrophies, but generally not after denervation. This suggests that the presence and maintenance of innervation may play a crucial role in the recovery of function. The precise mechanism(s) by which nerves regenerate and integrate with acinar cells in not known [16], and the impact of radiotherapy on this mechanism is less certain. Furthermore, it is not clear that the same mechanism is responsible in radiotherapy-induced disease.

Function alterations and xerostomia can be induced by medication, but the condition is usually transient and subsides when medication ceases [289]. The most common medications include anticholinergic, sympathomimetic, cytotoxic, and antimigraine drugs, muscle relaxants, opioids, retinoids, and cytokines [290]. Both dysfunction and hyperfunction are possible, and some drugs alter saliva chemistry to the extent that it becomes discolored [291].

Fractionated radiotherapy for head-and-neck cancers often results in dysfunction, even if the patient is unaware. Besides reduction in volume and flow rate, composition is impacted and an increased bacterial presence can result [258, 289]. In contrast to medication, radiation-induced xerostomia and dysfunction onset may be immediate, complications can be more severe due to severity of dysfunction and confounding complications (e.g., damage to nerves, vasculature, osteoporosis, or osteoradionecrosis), and can last indefinitely [26, 224]. Though the nature of the dysfunction is not precisely understood, its occurrence is well-known. Reports of recovery are more mixed and sometimes sporadic. Mossman et al. [292] remarked on the matter.

> *Results of studies of the time and extent of recovery of normal taste and salivary function in man following radiotherapy are contradictory. Several investigators have observed complete recovery of taste and salivary function in patients* 1-3 *months following treatment, whereas others have not. Although little or no improvement in salivary function has been observed in some patients at least two years following curative courses of radiotherapy, partial recovery* 8 *months after radiotherapy has also been reported.*

Eneroth et al. [293] reported the case of a patient irradiated with 65Gy (a dose that would normally cause complete dysfunction) whose parotid tissue remained functioning up to nine years following conventional radiotherapy. This remarkable but not altogether uncommon case suggests a recovery mechanism that is able to withstand intense radiation. It also highlights that we are not (yet) capable of predicting specific complications, but rather the *risk* of complications.

The mechanism of radiation damage in the parotid (which appears to be the most well-studied salivary organ in general) remains poorly understood. Studies have shown that the magnitude of salivary gland damage in mammals and humans increases in proportion with dose and irradiated volume [40, 229, 230, 231, 232]. Damage to the parenchyma must clearly impact normal function, but it is not clear what portion of loss is attributable to specific parenchymal tissues. What is known is what has reliably been observed: that acinar atrophy and chronic inflammation of the salivary glands occur when dysfunction or xerostomia are induced via radiotherapy [294].

It is thought that some dysfunction is attributable to loss of vasculature support. Vascular changes within the parotid resulting from irradiation have been observed, beginning generally with periarteritis (severe inflammation of the outer arterial coat) and endarteritis (severe inflammation of the inner arterial lining – the tunica intima). Inflammation progresses into fibrosis of the tunica intima and eventual destruction of the vessel lumena [295]. The slow progression of inflammation is generally thought to result from gradual necrosis rather than direct (acute) apoptosis [294]. Nearby bone is negatively impacted and can become brittle [295, 296]. While it is clear that loss of vasculature support can cause dysfunction, it is not known whether the fibrosis mechanism could itself support any form of functional recovery. The vasculature model, even if partially valid, therefore does not fully describe the interaction between radiation and dysfunction or xerostomia. Furthermore, even if the issue of recovery is ignored, the evidence that inflammation contributes substantially to dysfunction is weak [297].

Radiation damage to saliva-secreting acinar cells are also thought to contribute to dysfunction and xerostomia severity. Both acinar cell types

– serous and mucinous – atrophy to an extent that is obvious in routine histopathological observation [298]. However, saliva in patients suffering from the most severe xerostomia is generally thick, discolored, and contains an abnormally high level of mucins [237]. It is thought that the early response of acinar cells to lethal doses of radiation appears to be rapid atrophy without inflammation, which suggests direct radiation-induced apoptosis or selective damage to the plasma membrane may be the causative mechanism rather than necrosis [294, 299, 300, 301, 302, 303]. Radiation injuries are known to manifest within mere hours, even at low doses [298]. Inflammation from vasculature damage and acinar cell atrophy are therefore likely to be two different mechanisms that both contribute to dysfunction and xerostomia. Considering only observations of the change in saliva viscosity, serous cells appear to be less capable of surviving irradiation than mucinous cells. Parotid glands are comprised almost entirely of serous cells, and a majority of acinar cells in submandibulars are of the serous type. The loss of a significant fraction of serous cells due to radiological sensitivity in either glands would match with the dual observations of substantially reduced quantity of saliva with simultaneous thickening. This likely scenario could explain both dysfunction and xerostomia incidence. It could also explain the weak correlation between the two conditions if there was a small number of mucinous acinar cells within the parotid that varied in proportion with the serous cells from patient to patient. In patients with few parotid mucinous acinar cells, even a modest dose could destroy serous acinar cells, leaving the parotid essentially functionally inert. At the same time, the continued support from the relatively hardy mucinous acinar cells in the submandibular and accessory glands would result in thick, mucousy saliva. In this scenario, it is not clear what proportion of stimulated saliva would be attributable to parotid vs. submandibular glands, but presumably the organ with the greatest number of intact mucinous acinar cells would supply the greatest proportion.

Salivary gland volume reductions over the course of radiotherapy have been reported by many [74, 148, 297, 304, 305]. Both gross gland volume and submandibular acini size reduction can be induced via sympathetic denervation [306], but the ducts do not appear to substantially change. It is

thus unclear whether radiation damage to the nerve tissues directly contributes significantly. Rather, volume changes may merely be a symptom of acinar cell atrophy [297]. Volume reduction of parotid and submandibular glands can be manually induced in the absence of radiation damage through atrophy via ductal ligation and occlusion [297, 302]. Conversely, *de*-ligation or clearance of the occlusion is known to initiate an acinar cell regeneration process in which retained acinar cells not only proliferate, but also new formations of cells and parasympathetic innervation appear [302, 307, 308, 309, 310, 311, 312]. Salivary function is restored in this regenerative process [302, 307, 311]. This process also hints that broad recovery from radiotherapy-induced dysfunction may be possible. However, ligation and ductal occlusion likely involve very different atrophic pathways compared to radiation-induced atrophic pathways, and it is not yet clear if the same recovery can be induced in radiation-damaged glands. It is a promising line of inquiry though, so recent research efforts are described in section 8.2.

The above theories have not conclusively been demonstrated, though they are fairly basic first-order theories that do not stray far from observations. Many supplementary and alternative theories have been proposed to describe the mechanisms behind radiotherapy-induced dysfunction. It has been posited that: the presence of nitric oxide following radiotherapy [313], loss of salivary gland progenitor cells [314], or sublethal DNA damage which becomes lethal when acinar progenitor cells are undergoing reproduction [315] are mechanisms of functional loss. Such theories are covered in-depth in various review articles, including those of Konings et al. [303] and Carpenter and Cotroneo [297].

Given the evidence of multiple potential pathways to dysfunction (i.e., medication, autoimmune disease, radiation damage potentially from atrophy of acinar cells and/or inflammation, amongst others), the multiple pathways that could influence subjective xerostomia (i.e., the confounding effects of whole saliva and contributions from minor and accessory glands, saliva volume, flow rate, chemical composition, and mechanical properties), and observations of functional recovery by regeneration and reinnervation of parenchyma, it seems most likely that xerostomia/dysfunction are multifactorial and that no single mechanism is responsible. Rather, a multifactorial approach is likely

needed to accurately quantify toxicity risk, but it seems unlikely that a model could be concocted to incorporate all potentially confounding factors. In the following section, various factors and measurements (collectively 'facets') are selected to provide the greatest likelihood of detecting regional effects.

6.7 Overview of Toxicity Facets for Analysis

Several facets are recorded at the BCCA: stimulated and unstimulated whole-mouth salivary flow measurements and QoL questionnaire responses, each collected at four time points (before radiotherapy and each of three months, one year, and two years after therapy concludes). It is not feasible to perform a single analysis with so many response variables and predictors, even if non-parametric methods are employed. First, there is multicollinearity amongst predictors which may confound analysis. Second, combining some facets may lead to spurious findings (e.g., mixing stimulated and unstimulated facets from different time points could lead to physically nonsensical conclusions). Finally, including higher-order effects may result in high sensitivity to measurement noise that will thwart analysis. We therefore prioritize available facets for analysis. The relative merit of each selection is discussed in the remainder of this section.

6.7.1 Xerostomia or Dysfunction?

Xerostomia is the single most common complication following radiotherapy for head-and-neck cancers [10]. As a subjective condition, xerostomia questionnaire responses provide more detailed information about patient impact than objective flow measurements. Indeed, it is well known that xerostomia can have a substantial negative impact on a patient's social, psychological, and overall well-being [316]. In this sense, objective measurements do not have any bearing on xerostomia; if a patient with substantial flow *feels* like they have dry mouth, then many (psychological, social, and QoL) negative ramifications will manifest. Conversely, if a hardy patient with little flow does not notice a dry mouth sensation, then psychological, social, and QoL negative ramifications will not manifest. On the other hand, oral hygiene will

still be impacted.

Meirovitz et al. [317] advocates the use of patient-reported xerostomia over direct measurement. Others advocate the converse (e.g., [318]). Both facets are important. But there are trade-offs to using either (and both). Xerostomia is more relevant for psychological well-being, as patients with mild dysfunction may not even be aware of it. On the other hand, risk of dental caries and increased oral infection risk may increase regardless of patient awareness if dysfunction presents. Dysfunction is not subject to patient psychology and is likely more regular from centre to centre, especially if different xerostomia instruments are used. Conversely, objective measurements (e.g., flow rates, saliva composition or viscosity) are more specific. Dysfunction is likely more applicable to general clinical practice, and objective salivary flow measurements were therefore prioritized over subjective xerostomia scores.

6.7.2 Early or Late?

Figures 6.1 and 6.2 show saliva measurements before and each of three months, one year, and two years after conclusion of treatment. Summarized stimulated measurements are shown in fig. 6.1 and resting (unstimulated) measurements are shown in fig. 6.2. The measurement process is described in section 7.2; in brief, patients salivate into a beaker over five minutes and the total volume is weighed.

The median (red dot) and mean (blue dot) are shown in both figs. 6.1 and 6.2. On average, patient salivary function drops to around 30-50% of their pre-radiotherapy function. The magnitude of recovery is much less than the initial loss of function, on average. Additionally, the differences between one year and two year distributions are small compared to either baseline or three month distributions. This suggests long-term function approximately stabilizes between three months and one year after treatment concludes.

Another striking feature of figs. 6.1 and 6.2 is that dysfunction can (and often does) remain *two years after treatment or longer*. While the early loss of function (three months) is unfortunate, it appears to partially subside, one average, within the following nine months. Late dysfunction, on the other

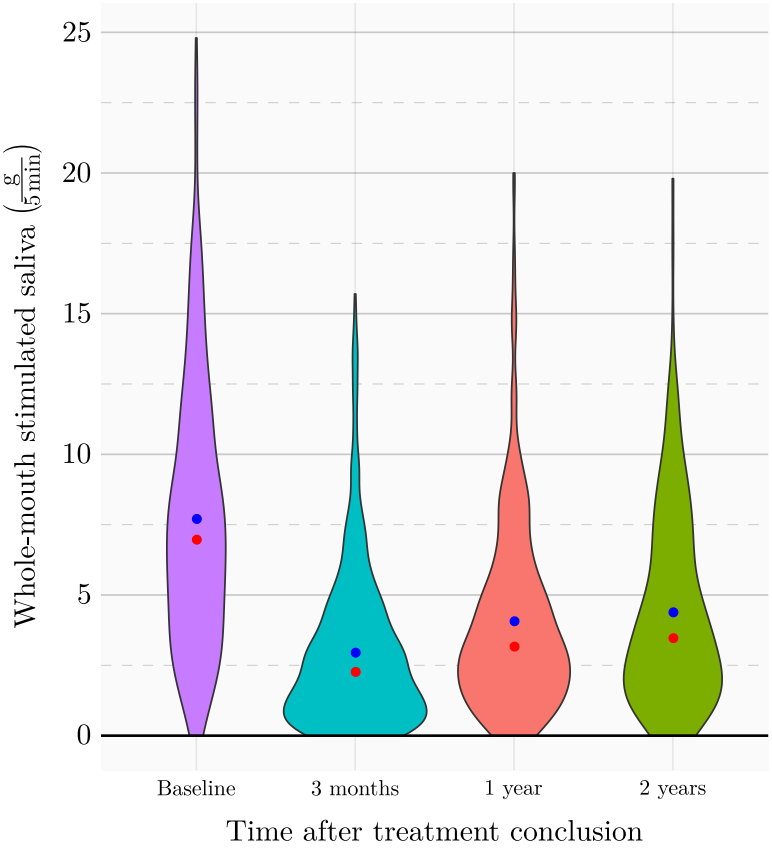


Figure 6.1: Violin plot of whole-mouth **stimulated** saliva measurements over time. The red dot represents the median, the blue dot represents the mean, and the shape represents a kernel density that estimates the measurement probability density. The optimal kernel density bandwidth was estimated by the method of [8]. Note the similarity of one- and two-year distributions compared with baseline and three-month distributions.

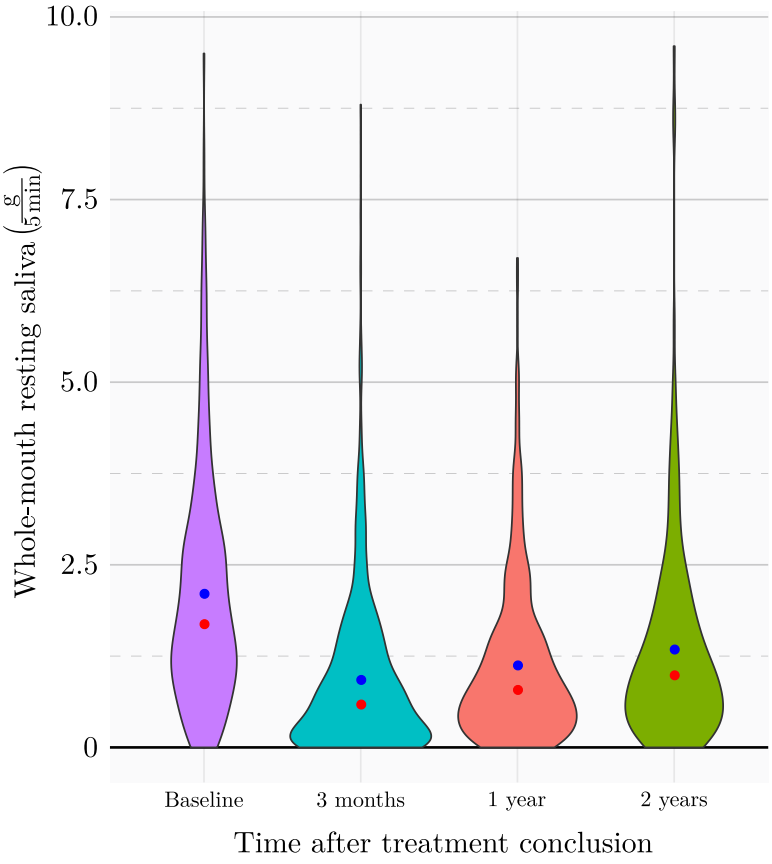


Figure 6.2: Violin plot of whole-mouth **resting** saliva measurements over time. The red dot represents the median, the blue dot represents the mean, and the shape represents a kernel density that estimates the measurement probability density. The optimal kernel density bandwidth was estimated by the method of [8]. One- and two-year distributions are differentiated than the stimulated case, but are still the most similar compared with baseline and three-month distributions.

hand, can last for years or even indefinitely. Therefore late salivary function presents an overall greater burden on survivors and was prioritized over early dysfunction.

6.7.3 Loss or Recovery?

The 'early vs. late' debate implicitly focused on absolute measurements rather than normalized measurements. An alternative approach, one that may appear at first to be more specific, is to focus on the difference of measurements. The difference between baseline and three months, before substantial recovery has taken place, could be called 'early loss.' Likewise, 'recovery' could be taken to be the difference of one or two year and three month measurements. This approach was taken during the author's MSc and resulted in numerous difficulties, which are described in the remainder of this section. Briefly, the core difficulty of this approach is magnification of noise and multiplicative baseline normalization emerges as the best simple normalization approach.

In a comprehensive study of the variability of salivary flow, Burlage et al. [319] noted that standard deviations ∼24% should be expected for measurement of whole-mouth stimulated flow. A large single study with 1427 healthy volunteers of varying sociodemographic backgrounds found 44% [320]. Blanco et al. [26] performed a repeatability experiment using five healthy volunteers. The baseline-equivalent measurement exhibited a standard deviation of only 27%, but suggested *patient* measurement variability may be higher due to disease, comorbid conditions, or confounding factors. There are many factors that could confound whole mouth salivary measurements, such as changes in body chemistry, hydration levels, and circadian rhythms [25]. Even mobile phone use impacts saliva production [61]. One consistently different factor between patient and volunteer studies is participant age. Though it is known that stimulated saliva production is not strongly associated directly with age, unstimulated production generally decreases and factors that themselves correlate with age, such as regular medication use, do negatively impact production [321]. Therefore, measurement noise may be greater than gener-

ally reported in volunteer studies. Most incidental reports of measurement variability indicate standard deviations around 30-40% of baseline which is consistent with age- or disease-related increases.

Stimulated salivary flow is reasonably stable over the span of two hours [25]. Pooling of repeat measurements to reduce variability quickly runs into issues, however. Continual saliva production cannot continue indefinitely, and since daily salivary output is estimated to be around 1.0-1.5L, dehydration at a level significant enough to impact repeated (i.e., back-to-back) flow rates could result [34, 322]. Given this limitation and other limits of clinical acceptability (e.g., clinical time), repeat measurements appear to have the capacity to make a negligible effect in reduction of noise [319]. It appears the measurement variability problem for whole-mouth saliva is indefeasible and therefore must be accounted for in the analysis design.

Examining absolute measurements conceptually involves a single measurement uncertainty[4]. Computing the difference of two measurements requires both uncertainties to be accounted for. Depending on the underlying process, the combination procedure (for two measurements W_A and W_B combined as $W = W_A - W_B$) can be

$$\delta W = \sqrt{(\delta W_A)^2 + (\delta W_B)^2}$$

or, more generally, when the nature of the uncertainties is unknown,

$$\delta W \leq \delta W_A + \delta W_B$$

where δW represents the uncertainty in W [323]. Thus, the ratio of uncertainty to signal is

$$\frac{\delta W}{|W|} = \frac{\sqrt{(\delta W_A)^2 + (\delta W_B)^2}}{|W_A - W_B|}$$

[4]Depending on the the normalization scheme employed. This is discussed shortly in section 6.7.6.

or more generally

$$\frac{\delta W}{|W|} \leq \frac{\delta W_A + \delta W_B}{|W_A - W_B|} \tag{6.1}$$

which, in both cases, grow large when the magnitude of W_A and W_B are both small compared to $\sup(\delta W_A, \delta W_B)$, even if $|W_A - W_B|$ is reasonably large. In all physically sensible cases it follows that

$$\frac{\delta W}{|W|} \leq \frac{\delta W_A}{W_A}.$$

Thus, the net effect of measurement differencing is a magnification of uncertainty compared to the absolute approach. Many patients at three months can produce very little (or zero) saliva, so the magnitude of the uncertainty (δW_{3m}) is large compared with the measurement (W_{3m}). When a difference is computed, δW_{3m} 'taints' $|W_b - W_{3m}|$ and produces relatively large uncertainties which reflect the low reliability of the difference. Using stimulated salivary measurements, population-derived medians, and assuming an intrinsic uncertainty of 35%, $\delta W/|W| \in [0.73, 0.97]$ where $W = W_b - W_{1y}$ depending on whether uncertainties were known to be independent and random. In contrast, $\delta W_{1y}/W_{1y}$ is 0.35 and if W_{1y} is baseline normalized (discussed below) then $\delta W/|W| \in [0.50, 0.70]$ – which is 39-47% smaller than the difference approach. The unstimulated case is worse because measurement magnitudes are all smaller, and the problem persists when W_{3m}, W_{1y}, or W_{2y} are used or when a small constant uncertainty factor is included (e.g., $\pm$0.1g/5min to account for scale calibration errors). Recovery presents the worst uncertainty ratios at $\delta W/|W| \in [1.6, 2.2]$ so that the magnitude of uncertainties even exceeds the signal magnitude.

Disregarding the uncertainty magnification issue, and despite interesting recent developments linking stem cells to recovery (mentioned earlier in section 6.6 and discussed in greater detail in section 8.2), it is not clear if recovery (currently) is more clinically relevant than loss. As noted in figs. 6.1 and 6.2, the magnitude of recovery is much smaller than the magnitude of loss. Earlier work on a portion of the BCCA data set showed the magnitude

of recovery had virtually zero dependence on the gross dose profile, with each patient recovering around 20-30% of baseline function regardless of dose, sometimes (rarely) even beyond baseline levels [324]. The finding is difficult to interpret. It may be that use of current clinical guidelines (based on mean doses) limits recovery. It is also possible that clinical guidelines are incidental, and that recovery truly does not have dose-profile dependence. What is clear, however, is that the prospects are limited for investigation of recovery using only clinical treatment plans that are subject to uniform parotid gland sparing procedures. Such investigations would be best done through radiobiological studies or clinical trials. On the other hand, given that modern treatment technologies are *currently* capable of delivering more specific dose-profiles with sub-organ features, and the *current* clinical magnitude of loss is greater than that of recovery, it therefore seems clinically prudent to investigate function loss rather than recovery.

6.7.4 Stimulated Saliva or Resting Saliva?

Resting saliva may play a greater role in regulating oral health than stimulated saliva, and submandibulars – which provide the majority of resting saliva – are known to be important for predicting xerostomia [233, 325, 326]. However, scenarios where salivary dysfunction have the greatest negative impact are predominantly those involving stimulation, such as eating, chewing, swallowing, and speaking. As seen in tables 6.1 and 6.2, baseline-normalized whole stimulated saliva measurements generally correlate more strongly with QoL questionnaires, or at least do not seem to correlate any worse than whole resting saliva. (The average correlation for stimulated saliva was 0.217 vs. 0.196 for resting saliva.) This finding is not unique to the BCCA cohort [227, 327]. Furthermore, stimulated saliva measurements are also greater in magnitude than resting saliva in all cases, on average, as can be seen in figs. 6.1 and 6.2. Use of more voluminous measurements will improve analysis precision since fewer of the worst affected patients will present sub-threshold whole saliva.

Therefore, stimulated saliva will be prioritized. Note that parotids are in

all analyses paired with stimulated saliva whereas submandibulars are paired with resting saliva owing to the relative contribution by each gland.

6.7.5 Parotids, Submandibulars, or minor glands?

As discussed in section 6.1, parotid glands are known to contribute the largest proportion of stimulated whole-mouth saliva. Parotids also form the basis for current clinical guidelines [236]. Parotids should therefore be the first-order organ considered for analyses investigating whole-mouth stimulated saliva dysfunction.

There are subtle differences in clinical contraints reported throughout the literature that complicate assessing non-parotid salivary gland importance. Blanco et al. [26] attempted to spare neither oral cavity nor submandibular glands. Rather, patients for which glands could be spared were excluded from their study. It my be unsurprising then that they found only a small association between xerostomia and non-parotid salivary gland dose. Similarly, Chao et al. [229] intentionally considered only high submandibular mean dose patients ($\geq$ 50Gy) which is likely to have impacted their (similar to Blanco et al. [26]) conclusions. Nishimura et al. [148] found no statistically significant mean dose dependence, though submandibular sparing was not explicitly described. Overall, submandibulars appear to be less studied in the literature compared to parotids.

As discussed in section 6.1, submandibular glands are known to contribute the greatest proportion of resting whole-mouth saliva. Their surgical removal and transplantation are known to impact xerostomia risk [233, 234, 235]. On the other hand, submandibular gland often sparing is somewhat controversial as they are known to frequently coincide with primary targets, and local control can be compromised [233, 326]. The submandibulars are smaller and less geographically distributed than parotids, which means sparing one is harder, and the capacity to separate effects from each submandibular will thus be harder than for parotids. Submandibulars should therefore be the second-order organ considered.

Regrettably, sublingual glands are not contoured at the BCCA and thus

cannot be directly included[5]. However, they are included in the broad oral cavity ROI. Previous studies are mixed: dose to the oral cavity, and thus to some accessory and minor salivary glands, has been found to be both significant [40] and insignificant [26] for prediction of xerostomia. Because the lesser glands contribute less saliva than larger organs, they should therefore be included in analyses only after parotids and submandibulars, and possibly after other accounting for other factors such as age, gender, dose-volume effects, and bath-and-shower effects (n.b. discussed in section 8.1).

Ideally, all salivary organs should be considered in analysis of regional effects. The reality is that contouring and other clinical support efforts are time consuming and must be prioritized. The parotid is most frequently contoured at the BCCA in order to adhere with clinical guidelines. Submandibulars are also frequently contoured, but less often. The oral cavity is often contoured, but least of all structures containing salivary glands, and the definition can be subjective. Also, patient positioning does not always include maintaining the shape of the oral cavity. Sublingual glands are rarely contoured. The availability of contours reflects their (current) clinical utility, which makes it difficult to investigate higher-order effects. Therefore, the parotid is primarily considered in this thesis. Submandibulars are investigated briefly, but lesser glands and the oral cavity are not considered.

6.7.6 Relative or Absolute?

Patient anatomy and physiology dictates the magnitude of saliva response, and what is considered low flow or highly viscous for one patient may not directly translate to other patients. In other words, salivary measurements are patient-specific. A brief discussion of salivary function measurement and xerostomia questionnaire response normalization was given in section 6.4. However, it did not address the issue of how to make the most of the data through normalization.

As shown in section 6.7.3, the most natural normalization method – subtraction of baseline – magnifies measurement uncertainties. The next most

[5]Note that studies involving sublingual glands are comparatively rare in the literature and often seem to be lumped into the more generic oral cavity.

natural method, division of baseline, yields a reasonable compromise between measurement uncertainty amplification and per-patient normalization. More precisely, with $W = W_A/W_b$, propagation of uncertainties for baseline-normalization yields

$$\frac{\delta W}{|W|} \leq \frac{\delta W_A}{|W_A|} + \frac{\delta W_b}{|W_b|} \tag{6.2}$$

generally, or

$$\frac{\delta W}{|W|} = \sqrt{\left(\frac{\delta W_A}{|W_A|}\right)^2 + \left(\frac{\delta W_b}{|W_b|}\right)^2}$$

when errors are known to be independent and random [323]. As shown in section 6.7.3, there is an uncertainty penalty incurred by baseline normalization compared to direct use of raw measurements (i.e., 0.35 vs. $[0.50, 0.70]$ in the example of section 6.7.3), but it is substantially less than the differencing approach (i.e., $[0.73, 0.97]$ or $[1.6, 2.2]$ depending on the signal of interest).

It is natural to wonder if there is a better means of bolstering signal without also amplifying uncertainty. In terms of δW and W, this question is tantamount to minimizing $\delta W/|W|$ for some physically meaningful function W involving at least two of W_b, W_{3m}, W_{1y}, or W_{2y}. First, it is more general and convenient to avoid assumptions about the nature of the measurement uncertainties especially since the scedasticity will change for arbitrary W. Uncertainties can be propagated in generality via

$$\frac{\delta W}{|W|} \leq \frac{1}{|W|} \sum_{i=1}^{N} \left|\frac{\partial W}{\partial x_i}\right| \delta x_i$$

where $x \equiv \{W_b, W_{3m}, W_{1y}, W_{2y}\}$ and $N \in [2, 4]$ for some function for some function $W(x)$ when $|W| \neq 0$ [323]. Any W involving only products or divisions of any two x_i trivially reduces to the RHS of eq. (6.2); any functions involving more than two reduce to the RHS of eq. (6.2) with an additional (positive) factor, and are thus strongly amplify uncertainties. Similarly, any W involving two or more additions or subtractions reduces to the RHS of

eq. (6.1), possibly with an additional positive factor, which makes $\delta W/|W|$ trivially larger than the RHS of eq. (6.2). So there are no simple arithmetical combinations of factors, including linear transformations, that improve upon multiplicative baseline-normalization in their ability to suppress measurement uncertainties. Clinical relevancy requires W to be monotonic in response to W_{3m}, W_{1y}, and W_{2y}. Non-linear transformations will generally skew uncertainties, and logarithm and power transformations are sometimes (controversially) employed to de-skew distributions to improve symmetry [328, 329]. Power transformations have already been shown to magnify uncertainties. However, logarithms may or may not magnify uncertainties. If $F \equiv \log W$ then it can be shown that

$$\frac{\delta F}{|F|} \leq \frac{1}{|W \log W|} \sum_{i=1}^{N} \left| \frac{\partial W}{\partial x_i} \right| \delta x_i$$

which may or may not be a stronger bound than that of $\frac{\delta W}{|W|}$ due to the factor $|1/\log W|$. For example, the factor will cause uncertainty amplification if $W = W_{1y}/W_b$ and population-average values are used. Perhaps more importantly, uncertainties will not *uniformly* by magnified, which could confound some analysis and may require delicate accounting of the nature of measurement uncertainties. If this is not possible linear transformations should be preferred, and the multiplicative baseline-normalization is therefore likely the most consistently optimal method. For purposes of analysis of regional effects, objective salivary measurements are multiplicatively baseline-normalized, yielding a value nominally within $[0, 1]$.

Xerostomia questionnaire responses may also require normalization, but are much more difficult to analyze. First, as discussed in section 6.4, the sample-population baseline score medians are all zero and therefore do not admit a reasonable multiplicative per-patient normalization. On the other hand, there is already a natural, common scale that each patient is subjected to, which presents a natural normalization opportunity. Second, there are many questions to consider, and some may be impacted by question order or adjacency, or possibly even recent (and distant) patient memories and mood.

Finally, though the questions are qualified, individual questions designed to probe specific aspects of saliva (e.g., the mucin content) may inadvertently amalgamate unrelated circumstances that confuse the specificity (e.g., confusing speaking and sleeping). Combining responses may likewise destroy specificity. Trying to select related questions results in even more subjectivity, but examining responses individually will reduce overall specificity and will necessarily impact the statistical approach to account for the multiple comparison problem.

Xerostomia question responses were therefore individually examined without additional normalization, but with appropriate multiple comparison corrections where appropriate.

6.7.7 Summary and Conclusions

Objective dysfunction is less susceptible to clinical variation, patient psychology, and easier to normalize. Late toxicity has a greater total impact on patients, while early toxicity is usually limited in duration and the potential for improvement is confounded by factors such as medication. Parotids and, to a lesser extent, submandibulars are the most relevant organs for analyses involving whole-mouth saliva. Relative normalization, when possible, provide a way to compare different individuals in a consistent way.

In this thesis, late, objective stimulated saliva was analyzed along with parotid dosimetric data. A follow-up analysis using late, objective resting saliva and submandibulars is also provided. In particular, the oral cavity and early measurements were not considered in any analysis, but all remaining available data was analyzed in some fashion.

Chapter 7

Toxicity Assessment: Instruments and Protocols

7.1 Clinical Xerostomia Assessment

An instrument is needed for self-assessment of xerostomia. At the BCCA a single-page hard-copy questionnaire is employed. It has been approved by the joint BCCA-UBC Research Ethics Board (REB) and patients must consent to participate. Patients responses are collected four times over roughly 2.5y (prior to radiotherapy and three times after radiotherapy is concluded: three months, one year, and two years, nominally). Patients answer each of the nine English language questions themselves. Verbal translation into Mandarin, Hindi, and Farsi is provided by bilingual nurses when available, but the majority of patients are able to read the questionnaire themselves.

All responses are interval-type responses that can range from zero (negative; meaning "not at all," "no," or "infrequently") to ten (affirmative; meaning "strongly," "yes," or "frequently"). All questions are worded such that a response of ten will affirmatively convey symptoms of xerostomia. The questions are (verbatim) as follows:

1. Rate the discomfort of our dentures due to dryness (if you do not wear dentures please check □).

2. Rate the difficulty you experience in speaking due to dryness of your mouth and tongue.

3. Rate the difficulty you experience in chewing food due to dryness.

4. Rate the difficulty you experience in swallowing food due to dryness.

5. Rate the dryness your mouth feels when eating a meal.

6. Rate the dryness in your mouth while *not* eating or chewing.

7. Rate the frequency in sipping liquids to aid in swallowing food.

8. Rate the frequency of fluid intake required for oral comfort when *not* eating.

9. Rate the frequency of sleeping problems due to dryness.

While all questions correlate with salivary measurements only weakly (see tables 6.1 and 6.2), questions 5 and 6 correlate consistently in the top-ranks, and in many cases are the top-ranked correlates. This may be due to questions 5 and 6 explicitly and directly polling oral dryness. Questions 4 and 7 may be specifically more sensitive to dysphagia (i.e., persistent difficulties swallowing) than xerostomia, but are certain to be sensitive to xerostomia too. Conflation of symptoms is known to commonly occur, and may further confuse the association between xerostomia and salivary dysfunction [330].

The BCCA questionnaire derives from the eight-question xerostomia-specific QoL questionnaire of Eisbruch et al. [40], but some phrasing was modified and an additional question was added (#1) with the aim of assessing dental practices pertaining to dentures. Evaluation of the reproducibility and validity of the original questionnaire was performed "*... according to standard methods of evaluating QoL instruments*" [40]. There are a plethora of self-reporting instruments that are xerostomia-specific or are QoL-focused but encompass aspects of xerostomia.

Ojo et al. [331] give a systematic review of (>50) QoL instruments relating to head-and-neck cancers. The aim was to assess the heterogeneity of questionnaire use in the literature. The most widely used xerostomia-related

instruments were the 'Dysphagia Inventory' of MD Anderson [332], the questionnaire of Eisbruch et al. [40], and the 'Oral Health Impact Profile' (14-item version) [333]. A significant basis for the selection of an appropriate BCCA questionnaire was minimal respondent and administrative burden, since it would be administered several thousand times; the shortest xerostomia-specific questionnaire was that of Eisbruch et al. [40]. While the unmodified original questionnaire of Eisbruch et al. [40] was found to be used in only 14 of the 710 considered studies, it appears to be the most prevalent xerostomia-specific questionnaire in use.

It is worth reiterating that, whichever instrument is used, most studies report a rapid decrease in patient QoL shortly after radiotherapy with gradual improvement over the following year. It is unclear precisely how this observation relates to salivary dysfunction, or is impacted by patient perceptions of oral dryness, or the emotional burden of cancer and cancer treatment. Additionally, given the large number of assessment instruments and especially their variations, objective flow presents as a more universal measure.

7.2 Clinical Salivary Function Assessment

A comprehensive and unusually thorough[1] study by Navazesh et al. [334] found that unstimulated whole saliva and stimulation (chewing without tasting) had the greatest test-re-test repeatability. This is the procedure employed at the BCCA for objective saliva measurements. The procedure is widely used and has been described numerous times in the literature (e.g., [26, 229, 335]). While it appears to have been well known even prior to 1992, the specifics of the procedure were adapted from the report from Chao et al. [229] and are described here.

Assessment immediately followed administration of the xerostomia questionnaire from section 7.1, namely prior to radiotherapy and three times after treatment concluded: three months, one year, and two years, nominally.

[1] Practice saliva collection was performed and the chew rate was controlled via synchronization with a metronome.

Patients were asked to avoid consuming food one hour prior to assessment. They were asked to fill out the questionnaire while resting. Excess saliva was cleared from the mouth, and the patient was asked to lean forward and expectorate whole-mouth (unstimulated) saliva into a pre-weighed collection cup. After five minutes, the cup was weighed. Stimulated whole-mouth saliva followed using the same procedure, but a small block of flavourless paraffin wax was provided to chew on. Both procedures were supervised by qualified personnel to ensure adherence.

Collection of time-integrated saliva volumes allows estimation not only of total volume but also sustained flow rate. Instantaneous (a.k.a. 'on demand') or nearly-instantaneous flow may be the most perceptible aspect of stimulated saliva. For example, chronically being unable to swallow the first few bites of a meal, or failing to develop a full mouth salivary response when a tart substance is tasted are likely to invoke strong negative emotional responses. However, as previously discussed, saliva measurements are notoriously noisy and estimation of the instantaneous flow rate may involve only <10mg of saliva per second. Estimation by whole-mouth collection is therefore not feasible.

Whole-mouth saliva is contributed from both major and minor salivary glands. Organs with left-right symmetry separately contribute. Normal anatomy, disease, damage (e.g., trauma, radiation, medicinal), or other factors may lead to an unbalanced contribution from individual organs, and thus whole-mouth saliva measurements convolute the signal from each. There are well-known suction apparatuses that can measure contributions individual glands, including *Carlson-Crittenden collectors*, which were first described over a century ago [336]; the *Lashley cup* modifications and variations thereof [337, 338]; and *Schneyer's devices* [339, 340]. Compared with whole-mouth saliva measurements, dose response discrepancies have been noted for individual parotid saliva measurements [40, 229]. It is not known to what extent such discrepancies result from differences in collection technique [341]. It remains unclear if individual parotid flow measurement is more or less relevant than whole-mouth saliva. Regardless, minimization of clinical impact was the primary motivation for collection of the decidedly low-technology

whole-mouth saliva rather than use of specialized devices. Other whole-mouth collection methods have been devised, such as the Wolff method, based on the weight loss of a hard candy over a fixed duration, and manual oral swabbing. Both have been shown to be less repeatable than the method employed at the BCCA [334, 342].

Due to clinical constraints, it was not possible to ensure repeat measurements occurred at a consistent time of day. Collection predominantly occurred between 9:00 and 15:00.

7.3 Conclusions

BCCA xerostomia and dysfunction assessment protocols are based on the most reliable saliva collection techniques and the most widely-used xerostomia-specific QoL instruments.

Chapter 8

Other Relevant Topics From the Literature

8.1 Are Parotids Serial or Parallel Organs?

Parotids have been treated as parallel organs since the inception of radiotherapy, and current practice makes use of mean dose to the parotids. This is partly because the distribution of functional burden appears to be homogeneous *to first order*, and partly out of convenience to work around limitations of early treatment devices [343, 344, 345]. Dose-volume histograms are also used for treatment planning purposes; their validity is predicated on the assumption that organs are parallel and that functional burden distribution is homogeneous. Modern radiation delivery techniques have shed new light on the question of whether parotids are parallel or serial – or, rather, whether there are potential clinical gains to had by accounting for deviations from the purely parallel approximation. In the remainder of this section a brief overview of recent developments is given.

In 2005, Blanco et al. [26] considered six distinct whole mouth saliva models motivated by physical arguments. All involved exponential suppression of function with increased radiation dose. The best-performing model was a parallel-exponential model that treated the parotid as if it were composed of independent functional subunits. Belief in the applicability of the model

appears to have been limited, though, because simple models were ultimately advocated [26]. Many others have concluded the same (e.g., [346, 347]). On the other hand, the conclusions of this 'early modern' re-examination appear to have reinforced the belief that the parotid is parallel, which more or less continues even now.

Many explicit dose-volume effects have been reported for parotid glands. For example, Konings et al. [348] (and later Konings et al. [349]) reported late dose-volume effects pertaining to salivary dysfunction in the parotid glands of rats. The greater the partial volume of parotid that was irradiated, the greater the magnitude of dysfunction that resulted.

In 2010 Houweling et al. [347] pooled data from multiple centres for a large study involving 384 patients. Their aim was to select the best Normal Tissue Complication Probability (NTCP) model amongst the Lyman-Kutcher-Burman, mean dose, relative seriality, critical [sub-]volume, parallel functional subunit, and dose-threshold models. No model could be rejected, but the mean dose model was recommended. The same year Deasy et al. [236] reviewed the literature on late parotid xerostomia dose-volume effects and devised clinical guidelines based on mean whole-parotid doses that would nominally avoid severe late xerostomia. The mean dose paradigm, which forms the basis for current clinical guidelines, is not purely dose-volume. Rather, because the statistical mean is sensitive to every datum, mean dose is sensitive to dose in all regions within a ROI. Nonetheless mean doses do not account for *specific* spatial information and are rooted deeply in the parallel organ paradigm.

By 2010 dose-volume effects had been decried for nearly 30 years. The convenient assumption that functional burden is homogeneously distributed throughout the salivary organs was, and remains, widely criticized by radiobiologists [350, 351, 352]. Dose-volume models are based on the assumption of homogeneity. Early reports used the Lyman-Kutcher-Burman model, which uses a power law with an exponent that controls the strength of the volume effect and is only applicable for single-organ contributions [346]. In 2000, Hopewell and Trott [353] wrote (about the homogeneous distribution theory in general) that

> *There is little or no [dose-]volume effect for structural radiation damage, however, some very pronounced [dose-]volume effects have been reported for functional damage. Volume, as such, is not the relevant criterion, since critical, radiosensitive structures are not homogeneously distributed within organs.* [...] *Volume effects in patients and experimental animals are more related to organ anatomy and organ physiology than to cellular radiobiology.*

It therefore seems that volume, from a radiobiological perspective, is merely a surrogate for discretely distributed parenchyma (or some other critical tissues) that are only *approximately* homogeneously distributed at the whole-organ scale. Over a decade ago, when Konings et al. [348] reported dose-volume effects in rat, they also found that the character of the dose-volume effect differed in cranial and caudal aspects. This is an example of a weak regional effect. However, this aspect of the finding appears to have been somewhat disregarded until recently.

In the latter part of the decade spanning 2000-2010, the introduction of IMRT led to improvements in treatment outcomes and QoL. For example, Mortensen et al. [283] showed that incidences of both xerostomia and dysphagia were reduced following the introduction of IMRT. Others have found the magnitude of QoL impairment lessened, and recovery was both more rapid and more complete compared to conventional radiotherapy [226, 241]. However, the improvements in treatment capabilities highlighted shortcomings of various dose response models, which became highly scrutinized. The failure of NTCP models in animals was investigated by van Luijk et al. [354]. Dijkema et al. [355], using ten years of patient data, reported failure of mean dose based models to fully describe the effects of radiotherapy on parotid glands.

Other deviations from a purely parallel organ were reported prior to development of the modern clinical guidelines. The most striking was reported nearly almost concurrently with the guidelines of Deasy et al. [236] in 2009, namely the bath-and-shower effect reported by van Luijk et al. [356] in rat parotid. Again, the cranial and caudal aspects were irradiated. It was found

that the sub-effect threshold dose of one aspect depends on irradiation of the other aspect. A confirmation of the effect was noted in human parotids using xerostomia as the response variable in 2012 [233].

Recently, more direct evidence of regional effects has emerged. In 2015 van Luijk et al. [357] reported isolation of a 'critical' region; irradiation of the critical region is supposed to strongly influence patient outcomes. The degree of influence and specific sensitivity is not known, but histopathological work indicates the region is dense with stem/progenitor cells. van Luijk et al. [357] reports that a clinical trial is underway to assess the impact of sparing the confined critical region. If the trial is successful in showing the region is critical, it still may not be able to quantify the relative importance of other (potentially critical) regions. Likewise, reports of lobe sparing are limited in their ability to impact clinical treatment planning because they present information only relating to the (binary) decision to spare only a single region [358]. The findings are still noteworthy and could potentially impact treatment planning however. Even if sparing is found to be ineffectual for reducing toxicity risk, these findings *still* represent potentially clinically-relevant cracks in the parallel organ theory. It is hoped that the regional effects demonstrated by the present work will also deliver a significant blow to the atomic-organ clinical practice.

8.2 The Growing Importance of Stem Cells

Recently, advances have been made in understanding the role of stem cells in recovery of both parotid and submandibular gland function after radiotherapy damage. A seminal work on the subject is provided by Coppes and Stokman [302] that builds on nearly a decade of investigatory work.

As described in section 6.6, stem cells appear to play a central role in the restructuring of glands that have had temporary ductal ligation. In short, stem cells can (almost) fully restore functional capabilities by proliferating surviving acinar cells along with recruitment of new cells and fresh innervation. Newly-generated ducts express proteins usually observed only during embryogenesis [359]

These findings have sparked efforts to selectively protect regions containing the ducts so that greater stem cell recruitment from protected, stem cell-dense regions can more rapidly and consistently replenish damaged regions (i.e., [356]; cf. section 8.1).

Recently, researchers have attempted to transplant stem cells into dysfunctional parotid glands following radiotherapy [214, 360, 361]. Additional recovery has been observed. These findings have sparked a flurry of research into the mechanisms and clinical applicability of cell transplantation, as well as efforts to locate suitable sources of cells. Stem cells have been observed in the ducts of mice [362]. Candidate stem/progenitor cells have been found in uninjured salivary glands [363], pancreas [364], and minor salivary glands of the lips [365]. There is thus some possibility that transplantation could enter clinical practice rather than just remain an academic curiosity.

However, stem cell therapy to correct radiotherapy-induced dysfunction is not *currently* clinically feasible. Reports of functional restoration in a xenotransplantation model of human salivary glands have recently been reported that suggest it may be reasonably possible in the future [366, 367]. Efforts are currently underway to construct replacement salivary gland 'neotissues' for permanent implantation [368].

As described in section 6.7.3, the focus of this thesis is on reduction of dysfunction rather than improving recovery, and stem cells appear to be most relevant for recovery. So the relevance of stem cells are confined to possible regional effects resulting from inhomogeneous cell distribution, but there is no analytical way to derive cell populations from the BCCA cohort. Stem/progenitor cells are thus largely incidental for the purposes of this work.

8.3 Relevant Clinical Factors

The utility of many clinically-relevant factors has been investigated in salivary output prediction models. The use of multivariate logistic regression and other model parameter selection techniques are widely reported. In this section a sampling of these findings are provided. In brief, dosimetric factors are universally found to be most relevant for risk of developing xerostomia

and salivary dysfunction following head-and-neck cancer radiotherapy.

Using multivariate analysis, Chao et al. [229] investigated correlation between salivary function and radiation dose to the parotid glands. They found that the rate reduction of stimulated salivary flow at six months post-radiotherapy was not significantly influenced by a patient's gender, age, tumour stage, radiation technique (IMRT vs. conventional radiotherapy), or adjuvant chemotherapy. A parotid dose derivative was found to be the sole significant factor for xerostomia.

A similar, more comprehensive analysis by Blanco et al. [26] involved patient age, gender, ethnicity, date of treatment start, treatment technique (IMRT alone vs. otherwise), treatment aim (definitive vs. postoperative radiotherapy), Karnofsky Performance Status (KPS), whether adjuvant chemotherapy was delivered, tumour stage, treatment duration, histopathologic features (squamous vs. other), tumour subsite, and a derivative of mean parotid dose. The mean dose-exponential model was the most independently significant factor, followed by the considerably less significant gender and KPS factors [26].

A meta-analysis of xerostomia incidence in the elderly found that non-institutionalized elders had only a small additional xerostomia risk compared to the general public (17-40% vs. 6-39%) [369]. This suggests only a weak dependence of xerostomia with age. Wider socio-economic factors are known to moderately impact overall oral function. However, the strongest impact results from the most extreme and least common conditions involving psychiatric disorders, heavy smoking habits, and low socio-economic status. Furthermore, the impact appears to be greatest on xerostomia rather than dysfunction [370, 371, 372, 373].

In a multivariate study, El Naqa et al. [346] demonstrated a consistent preference of variables when constrained to five factors (only): mean parotid dose, gender, KPS (the three most significant), and technique and treatment aim (the two of considerably lower significance than the previous three). El Naqa et al. [346] recommend using a simplified model with few factors.

Teshima et al. [305] found a correlation between decreased parotid gland volume and decreased saliva production following radiotherapy. They noted

that no such correlation existed between *total* volume and *total* salivary output following radiotherapy. Many others have reported similar findings (e.g., [74, 148, 374]).

Recently Buettner et al. [233] employed Bayesian multivariate logistic regression, considering the parotid volume, submandibular gland mean dose, surgical removal of the ipsilateral submandibular, gender, tumour site (hypopharynx or oropharynx), age, utilization of neo-adjuvant chemotherapy, and hypertension assessment reports. Dosimetric factors were most significant, though surgical submandibular gland removal was found to significantly increase xerostomia risk.

Another recent study by Lee et al. [375] making use of the robust bootstrap-based *least absolute shrinkage and selection operator* (LASSO) statistical method found that mean dose to the either parotid gland were the most significant predictors for moderate-to-severe xerostomia. Patient age, financial status, tumour stage, and education were variously found to important, but not consistently. Other factors considered included patient gender, (current) marriage status, smoking status, history of alcohol abuse, and family history of xerostomia, amongst others.

Consistently, dosimetric factors have been shown to be most relevant for prediction of xerostomia and dysfunction. It therefore seems acceptable to forgo clinical and socio-economic factors and control for surgical, medicinal, and historical factors.

8.4 Factors Affecting Availability of Data

Studies have shown that logistical factors impact treatment prescriptions, availability of treatment options, treatment quality, and patient compliance to, and participation in, post-treatment procedures [376, 377]. Olson et al. [378] found that breast cancer patients from rural BC communities presented with more advanced disease than their urban counterparts. Conversely, medical travel[1] is well-known in the literature (e.g., [379, 380]) and is pronounced

[1] n.b. patients seeking more timely treatment or expertise not locally available – not to be confused with 'medical tourism', which specifically refers to patients seeking lower-cost, typically cosmetic treatments

in Canada [381]. One oncologist in California estimated that in 2006 approximately 40% of his patients were from outside the United States and 30% were from outside California [382]. Travel is known to disrupt continuity of care and follow-ups [383]. In BC and other places with centralized medical agencies, medical travel for cancers are thought to be most pronounced with advanced and specialized cancers for which specialized expertise can be sought, or waiting times are perceived to be too long, but a lack of academic literature or even empirical evidence limits understanding of the practice [377, 384].

All of these factors can be broadly classified as socio-economic. Direct inclusion of socio-economic factors does not appear to improve outcomes prediction (see section 8.3). However, they may significantly impact the data available for analysis, which can lead to a statistical situation referred to as Missing Not At Random (MNAR). Failure to account for confounding factors or use of MNAR data can lead to incorrect conclusions [385]. Socioeconomic factors were not recorded in the BCCA head-and-neck cohort, and retrospective assessment was not feasible. Generalizability of results from the literature is not easy to estimate, given the varying degrees of living standards and access to treatment options around the world. Therefore, an overview of findings from the literature with varying applicability is provided to help estimate the impact of the confounding missing data phenomena.

Since the aim of this study is to characterize patient outcomes, patient survival does not need to be accounted for. In toxicities that are directly life-threatening, failure to account for mortality may occlude the most severe toxicities. Salivary dysfunction and xerostomia can become strongly detrimental to overall QoL, but cannot directly (and do not often indirectly) cause mortality. There are even treatment options for patients dealing with chronic infections stemming from dysfunction, such as artificial saliva, medications that induce hyperfunction, and even transplantation or relocation of salivary glands. However, there is still potential for systematic bias if patients with the worst (or least) severe toxicities are systematically not reported. Treatment efficacy and local control are therefore de-coupled from development of toxicities, at least in the first-order.

Hypothetical Biases

Cancer treatment can be a significant economic burden. Even in Canada, where healthcare is publicly funded, radiotherapy treatment fractionation, frequent trips for follow-ups and monitoring, and concurrent modalities (i.e., imaging, administration of chemotherapy) cause significant strain on patients [386, 387]. Even seemingly mundane psychological and economic factors such as availability of (or payment for) parking are known to negatively impact patient involvement [377, 388, 389]. It therefore seems reasonable that socio-economic factors that mutually correlate with wealth, education, or social status, along with factors that impact aspects of disease, such as primary tumour site or proliferation of disease, could impact data availability.

Smoking history is potentially one such factor. While smoking status/history is known to substantially influence risk of head-and-neck cancers, predictions of toxicity risk/severity and treatment efficacy are only moderately improved when smoking status/history are included [201, 390, 391, 392, 393, 394]. The impact of smoking on data availability, however, is presently unknown. Consider a hypothetical pathway in which data availability could be negatively impacted; a rural-dwelling 'chain smoker' who spends much of their income on cigarettes. Because they live in a rural community, they do not have regular access to medical services, or the services they do have access to do not have substantial experience identifying cancers, and they present with an advanced stage cancer of the upper aerodigestive tract. (Many patients with head-and-neck cancers present at a relatively advanced stage [206], so disease in this hypothetical case may be extremely advanced or late stage.) Such a person would be predisposed to excessive head-and-neck cancer risk [391, 395] and may either find it difficult to travel for post-therapy follow-ups for logistic or economic reasons [386], or may be otherwise disinterested [272] (e.g., existing poor dental hygiene due to lack of regular checkups and lack of interest in attending concurrent dental examinations). Radiotherapy treatment would likely need be aggressive to obtain local control, and extensive lymph node irradiation would ensue. The parotid would thus be heavily irradiated. Socio-economic forces in this hypothetical case would result in

(1) increased cancer risk, (2) advancement in tumour stage and reduced prognosis, and (3) self-censorship of study involvement.

Whether this specific hypothetical case is far-fetched is unknown. Furthermore, it is merely one of many such pathways to developing a systematic bias in the available data. Other hypothetical self-censorship differences stemming from socio-economic factors might include greater allotment of sick leave for white-collar workers, greater accumulated savings and therefore less sensitivity to economic hardship and/or reduced reliance on friends and family for the wealthy, and greater treatment options for those in populated urban areas [389].

Literature on the interplay between socio-economic factors and data availability is sparse, so reliable estimation of their impact is not currently possible in generality. However, specific factors have received focused attention. Relevancy to the BCCA are discussed for each.

Rurality

One comparatively well-studied factor is distance to treatment centres. It is known that distance to a clinic negatively impacts patient involvement, such as attendance at follow-up appointments [388, 389]. It is also known that logistic factors impact utilization of radiotherapy [272]. It is not clear, however, whether there is any link between toxicity risk or severity and distance to the nearest treatment centre within BC. The BCCA is province-wide and implements similar practices throughout the province, including clinical OAR sparing. Currently mean doses are used, which are not sensitive to specifics of the dose profile, and thus no intra-parotid sparing is performed at any site. The same planning software is used to generate plans, and there are provincial tumour groups set up to collectively review specific cases, meaning there is open communication between sites, experience sharing, and continuing education amongst staff. Recent unpublished work by Olson demonstrated that treatment outcomes were indistinct for head-and-neck cancer patients treated at different BCCA centres [396]. Rurality therefore does not seem likely to have had a significant impact on data availability.

Additionally, the BCCA head-and-neck cohort was selected predominantly from large urban centres (Vancouver and Fraser Valley) rather than from all five provincial centres, which helps to further minimize impact clinical variations.

Ethnic Factors

Patients of Asian ethnicity are more susceptible to Nasopharyngeal Carcinoma (NPC) than those of other ethnicities [13, 397]. The *Hong Kong NPC Study Group* proposed the now common 'maxillary swing' surgical method and are world-renounced for specializing in NPC [398, 399]. Because a significant number of patients in the BC lower mainland, and therefore in the patient cohort, are of Asian ethnicity, it is possible that some will engage in medical travel and seek specialized treatment elsewhere. However, NPC is considered rare with a world-wide incidence of <1 per 100k per year [400]. Ethnic factors are therefore believed to have a low impact on data availability. Furthermore, in the specific case of NPC, medical travel may potentially help regress the BC incidence rate toward global averages.

Competing Modalities

Potentially confounding factors that could bias data availability are not limited to socio-economic factors. For example, alternative treatment options. Chemotherapy is most often administered concurrently with radiotherapy [202], but not always. Studies involving novel or specialized chemotherapy agents appear to most often select patients based on genetic predisposition, or viral status rather than disease stage or proximity to OARs (except in exceptional cases where radiotherapy treatment complication risk is exceptional, such as immediate proximity to the spinal cord, or if patient survival risk is low using existing treatment options). However, many chemotherapy agents, even those routinely used, are known to induce transient xerostomia [225, 401, 402]. Addition of chemotherapy afforded an overall reduction of toxicity compared with conventional radiation-only therapy [202, 403]. To mitigate potential bias, in all analyses we have therefore excluded patients

receiving non-standard chemotherapy agents. In addition, use of late salivary function rather than early function will further avoid conflation of chronic sequelae with transient chemotherapy-induced xerostomia or dysfunction. Other transient effects such as temporary changes in diet, depression, sleeping habits, and other early factors that may directly or indirectly affect hydration and salivation are also minimized this way [322, 404, 405].

Surgery (both relating to the cancer or previous surgical intervention) is another treatment option that impacts salivary function and xerostomia. In all analyses we have excluded patients with prior interfering surgeries. Otherwise, assessment of surgically-induced dysfunction, potentially confounding effects, and estimation of the impact on data availability due to patient enrollment in surgical treatments is well beyond the scope of this thesis. However radiotherapy with concomitant chemotherapy has displaced surgery as the primary treatment modality and is widely used in the clinic [202, 403, 406, 407].

Conclusions

No comprehensive studies of the impact of missing data on radiotherapy analysis was found in the literature. Surprisingly, only a few specific remarks were found. It is therefore difficult to assess how missing data may impact analysis.

On the other hand, whatever the impact of socio-economic or other factors on data availability, the BC cohort is likely one of the most reliable and cohesive large head-and-neck cancer data sets available. The Canadian population is one of the most ethnically diverse 'western' countries [408] and a provincial mandate means that all radiotherapy services are delivered through the BCCA. Medical services are not privatized in Canada meaning that all patients are able to afford (at least out-of-pocket) treatment expense, and collaboration between individual centres is high. Furthermore, nearly a decade of head-and-neck cancer outcomes (all using the same assessment procedures) are available. The results of this study are therefore likely to be *at least* as valid as the majority of studies reported in the literature, especially

if multi-agency data sharing was employed.

8.5 Clinical Recommendations

In 1999, Eisbruch et al. [409] recommended a mean dose $\leq$26Gy in 30-35 fractions to permit substantial sparing of the parotid. In 2003 Amosson et al. [267], based on patient-reported QoL questionnaires, found that patients felt they had 'too little' saliva when the contralateral parotid received a mean dose $\geq$22.5Gy. Six years after Eisbruch et al. [409], in 2005, Blanco et al. [26] recommended a mean dose $\leq$25.8Gy. Variation in recommendations have since appeared to be minimal, though many have been given [273].

In 2010 a joint effort by many researchers, authors, reviewers, and support personnel provided comprehensive summaries of the available dose-volume and outcomes data and accompanying clinical recommendations referred to as the Quantitative Analysis of Normal Tissue Effects in the Clinic (QUANTEC) guidelines. For parotids, <20Gy mean dose to the contralateral gland or <25Gy mean dose to both is advised. It is currently recommended to follow the QUANTEC guidelines [236, 273, 410]. The BCCA adheres to these guidelines.

Chapter 9

Statement of Research Questions

9.1 Statement

Chapters 2 to 8 have presented relevant introductory topics that frame and motivate the research topic, but also permit the specific aims to be more clearly stated. The goals of this thesis are to address the following questions.

1. Are all regions of the parotid gland homogeneously responsible for radiotherapy-induced dysfunction? Will delivering dose of equal magnitude to different regions of the parotid lead to an appreciable difference in risk of dysfunction?

2. If the regional importance is not homogeneously distributed, are there aspects that are more important? Are they clustered (i.e., are there 'critical' regions)?

3. Can the relative importance of parotid aspects for dysfunction be quantified? Can the entire gland be quantified?

4. If a strong effect is found for parotid dysfunction, does it manifest in the same way for xerostomia?

5. If a strong effect is found for parotid dysfunction, can similar effects be demonstrated for submandibulars?

Elucidation of regional effects in parotid may lead to improvements in treatment planning and either reduced risk or reduced severity of radiotherapy-induced toxicities for head-and-neck cancer patients. Demonstration of regional effects will be sufficient to warrant further research, but isolation of critical regions and quantifying their importance relative to less critical aspects within the parotid will have greater (immediate) clinical relevance. Quantification of the relative importance of the *whole* parotid will have the greatest clinical relevance since then toxicity risk could be more completely integrated into the treatment planning process. Finally, if other regional effects can be found (i.e., involving xerostomia, resting saliva, or submandibular facets) then a viable alternative to the current clinical practice of treating salivary organs as atomic (i.e., irreducible) parallel organs may be realized.

9.2 Outline of Approach

The following chapters describe and implement an analysis that is capable of achieving the research goals.

A glut of information is available to analyze. In order to provide the most clinically-relevant and widely applicable assessment, the analysis should (at first)

- employ segmentation to deal with morphological deviations,
- consider dysfunction rather than xerostomia,
- prioritize parotids over other salivary organs,
- make use of stimulated saliva rather than resting saliva,
- make use of late rather than early patient outcomes data, and
- make use of patient-specific baseline-normalization.

Several questions remain. First, discussion of the specifics of segmentation have been deferred to chapters 10 and 11; chapter 10 describes the

segmentation approach in generality, whereas chapter 11 presents a specific approach that navigates around a major analytical pitfall and is particularly suited to assessment of regional effects.

With the segmentation procedure outlined, analysis of the link between parotid sub-segment mean dose and whole-mouth stimulated saliva measurements proceeds. Chapters 12 and 13 describe two different analytical approaches; chapter 12 describes a model-based (i.e., parametric) approach, and chapter 13 describes a model-free (i.e., 'non-parametric') approach. The parametric approach is limited in that it cannot reasonably scale to accommodate many sub-segments, but it provides a crucial, well-grounded link between importance and clinical relevance that is needed for finer analysis. Using this link, the non-parametric method described in chapter 13 is able to perform analysis on fine sub-segments – all the way down to $1/96^{\text{th}}$ of a parotid, each with a volume of $\sim 0.3\text{cm}^3$ on average.

Chapters 12 and 13 successfully demonstrate regional effects in parotid, so two addenda are provided. First, a limited follow-up analysis is performed on submandibulars and using xerostomia rather than dysfunction in chapter 14. Second, an MR imaging protocol is developed in chapter 15 that may be able to detect salivary gland parenchyma. Detection of parenchyma *before* radiotherapy could potentially provide *patient-specific* estimates of toxicity risk rather than reliance on population-averaged importance.

Overall conclusions and a summary of contributions is provided in chapter 16.

Part II

Analytics

Chapter 10

The Basics of Segmentation

10.1 Introduction

Segmentation is open-ended. It can be applied in any number of ways which strongly depend on the representation of the underlying geometry. In this chapter the basics of segmentation as it is applied to ROI are described. Enough information to re-create the specific segmentation approach from basic principles.

Examples of ROI contour segmentation were shown section 4.3 – this chapter provides a more generic look at the *approach* rather than specific examples. It also provides sufficient preparation for chapter 11, which compares the merits of two distinct approaches to handling $\mathbb{R}^2$ planar contours embedded in $\mathbb{R}^3$ when volumetric aspects of ROIs are desired.

10.2 `DICOMautomaton`

`DICOMautomaton` is a computational framework developed by the author that presents a comprehensive system for manipulation and segmentation of contours and ROIs. Capabilities of have been described elsewhere (i.e., [7, 411, 412]), however the mathematical approach has not been described and is therefore presented here. `DICOMautomaton` is capable of rapid and lossless segmentation by operating directly on ROIs (i.e., a 'vector' approach) rather

than rasterizing (a 'bitmap' approach; referred to as 'generating masks') or volume reconstruction. `DICOMautomaton` operates on simple geometric primitives, including points, lines, line segments, and planes. Segmentation consists of the composition of elementary geometrical problems, and can be implemented using any computational geometry kernel, such as Boost.Geometry [413] or the Computational Geometry Algorithms Library (CGAL) [414, 415]. The vector approach is presented in the remainder of this chapter.

10.3 Solutions to Elementary Geometrical Problems

Fast and reliable computational algorithms for elementary $\mathbb{R}^2$ geometry problems have been described elsewhere and are not repeated here. However, a list of the primitive operations necessary for a vector segmentation approach are provided along with references to efficient algorithms. The list is short because the geometric primitives (points, lines, line segments, and planes) can be represented using a compact set of basic, atomic entities: points and vectors.

To perform vector segmentation, the following routines need to be implemented.

- Compute the distance between a line and a point [416].
- Compute the distance between a plane and point. The *signed* distance, which inherits a negative when the point is on the negative side of the plane, is useful. It is often ambiguous which side of the plane should be positive (especially since the coordinate system is not common to all patients), so an orientation must be consistently enforced.
- Find the intersection of a plane or line segment with another line segment [417]. If only the finite field is considered, there is a degenerate case where the plane and line segment coincide over the entire line segment which must be handled correctly.
- Compute the area and 'slab volume' for simple planar polygons [418]. 'Slab volume' was defined in section 3.4.1; in brief, it is the volume of

a planar polygon extruded perpendicular to the plane of coincidence. *Signed* area, which takes a polygonal orientation to arbitrarily be positive, is useful to describe polygons with holes, but is not strictly necessary for segmentation. As with signed planar distance, the issue of ambiguous orientation must be dealt with because the polygons ($\mathbb{R}^2$) are embedded in $\mathbb{R}^3$.

- Find the 'centroid' of a polygon, which is equivalent to the centre-of-mass when polygons are simple (or weakly simple) and the interior is assumed to be filled with a uniform density material [419].

- Find the optimal planar orientation when the plane intersects a (single, known) point (e.g., the centroid). This optional routine allows contours to be de-coupled from images, which describe the planes each contours coincide with. It is especially helpful when image planes are oblique. A least-squares solution has a closed form that is trivial to solve for.

- Projection of non-planar points onto a plane. This routine permits $\mathbb{R}^3$-embedded contours to be projected into $\mathbb{R}^2$, which can simplify implementation of other routines, such as computation of area and location of centroids.

- Determining if a point is interior or exterior to a planar polygon[1] [420]. This routine is crucial for differentiating entities within ROIs from those exterior. Planar projection is needed here unless perpendicular extrusion is employed, but extrusion is more costly due to the greater dimensionality.

The composition of these primitive operations into segmentation are described in sections 10.4 to 10.6.

[1] This problem has many names and is known as the "is point in poly" problem in some computational literature.

10.4 Planar Segmentation

Planar segmentation involves cleaving ROIs using infinite planes. For the purposes of planar segmentation, individual contours are treated as though they are two-dimensional planar closed contours embedded in $\mathbb{R}^3$ (i.e., they have zero volume and zero thickness).

Planar segmentation is accomplished, in a nutshell, by specifying a cleaving plane, walking each contour to locate the intersecting line segments, determining the precise intersection point and creating a new internal vertex, connecting internal vertices in a manner congruent with the contour orientation, and sorting resultant sub-segments by the side of the plane on which they rest. See fig. 10.1 for a graphical depiction.

Note that this procedure is indifferent to ROI 'thickness.' However, any procedure that makes use of planar segmentation and is sensitive to volume needs to account for plane-contour obliquity when specifying the cleaving plane. Slab volume may over or under-estimate the true volume, however accounting for volumes explicitly makes the problem considerably more difficult, including requiring primitives with volumetric extent rather than embedded planar contours with only an implicit thickness. If a two-manifold surface is reconstructed from an ROI then Gauss's theorem can be adapted to discretely compute enclosed volume [418]. However, using slab volume leads to no worse sub-segment volume estimates than estimates of the original ROI volumes, avoids ROI tessellation and the inherent topological ambiguities, and imposes substantially lower computational burden compared to two- or three-manifold methods [7], so the difference can often simply be ignored. Mitigation techniques include aligning planes and contours (so there is no obliquity) and iterative estimation of volume using slab-volume or by sampling vertices of a regular grid.

10.5 Projective Segmentation

Planar segmentation is confined to produce sub-segments that have convex (planar) faces. While this improves efficiency when recursive segmentation is performed, reducing the number of vertices that need to be processed

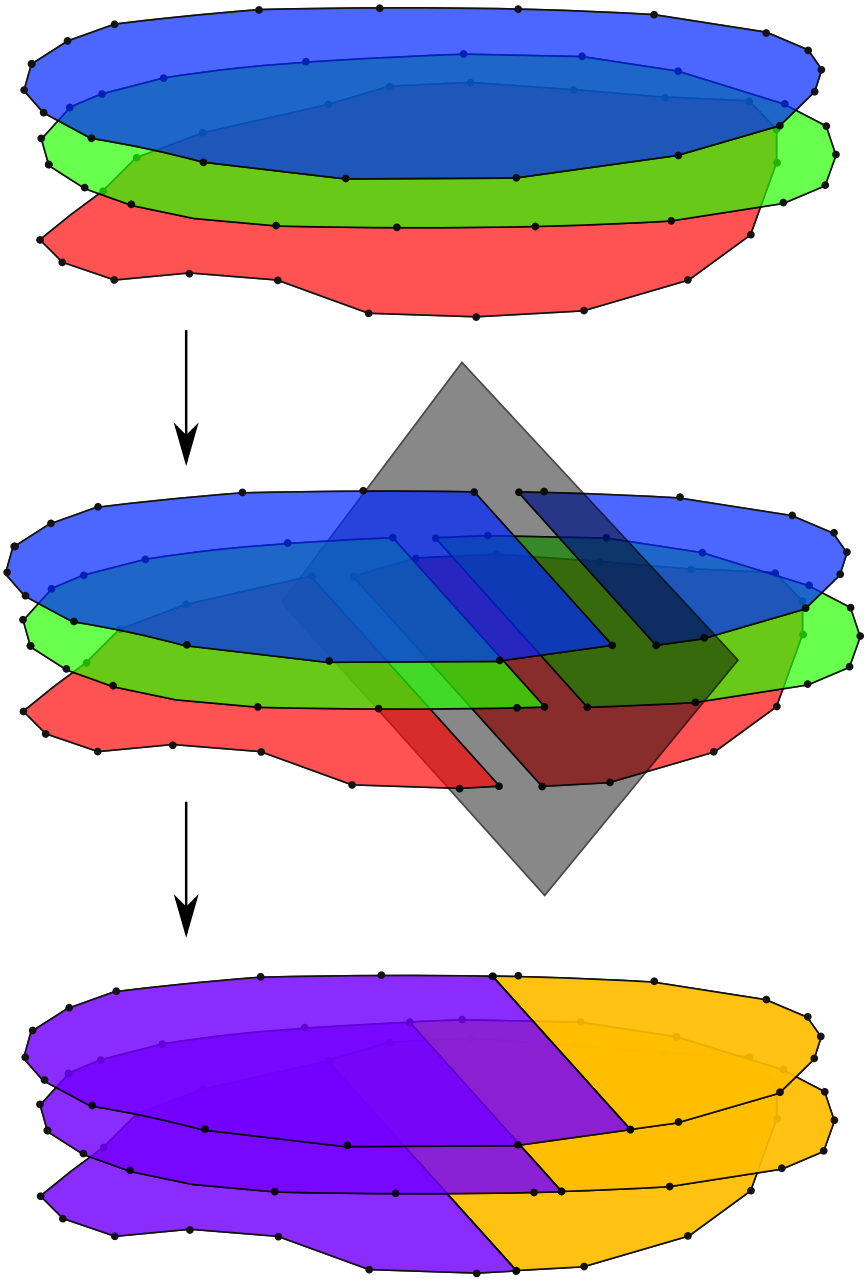


Figure 10.1: Demonstration of planar segmentation resulting in two sub-segments (collectively above and below the plane). Planar segmentation is always well-behaved when polygons are simple or weakly simple (i.e., holes with partial seams).

(to a single vertex per face for sub-segments not containing original ROI surface in as few as six segmentations), it limits the shapes representable by sub-segments.

An alternative technique is called projective or ray-casting segmentation, in which contours are segmented by specifying a casting direction and fractional width, sending rays through the contour at each vertex, backtracking until the fractional width is located, and creating a new internal vertex. Internal vertices are strung together in a manner congruent with the contour orientation and resultant sub-segments are sorted by orientation along the internal edge. See fig. 10.2.

Projective segmentation, which in some sense can more closely approximate what is implied by 'splitting a volume in two' by incorporating surface details, is more strongly affected by contour minutiae. Similar to difference between the mean and median discussed in section 4.3, projective segmentation is more sensitive to contour shape than planar segmentation and can produce sharp 'jumps' along the seam, particularly at the seam extrema (see fig. 10.3). Additionally, while cleave vertices are confined wholly within the original contour, the cleave edges are not (also shown in fig. 10.3) which can result in sub-segments with greater or lesser combined area/volume than the whole. On other words, area/volume is not strictly conserved. These failures are not pathological, and can occur whenever positive curvature is surrounded by negative curvature and dis-aligned with the casting direction.

Alternative approaches involving ray casting, such as spacing out rays evenly (or more finely such as recursively at vertex midpoints) over the polygon diameter only partially mitigate failures. In some cases they can make the problems worse. Additionally, projective segmentation can only be applied to individual contours. Combining adjacent sub-segments can lead to inconsistent seams and jumps (see fig. 4.10). Therefore, though it is sometimes more intuitive than planar segmentation, recursive projective segmentation is not reliable enough for assessing regional effects.

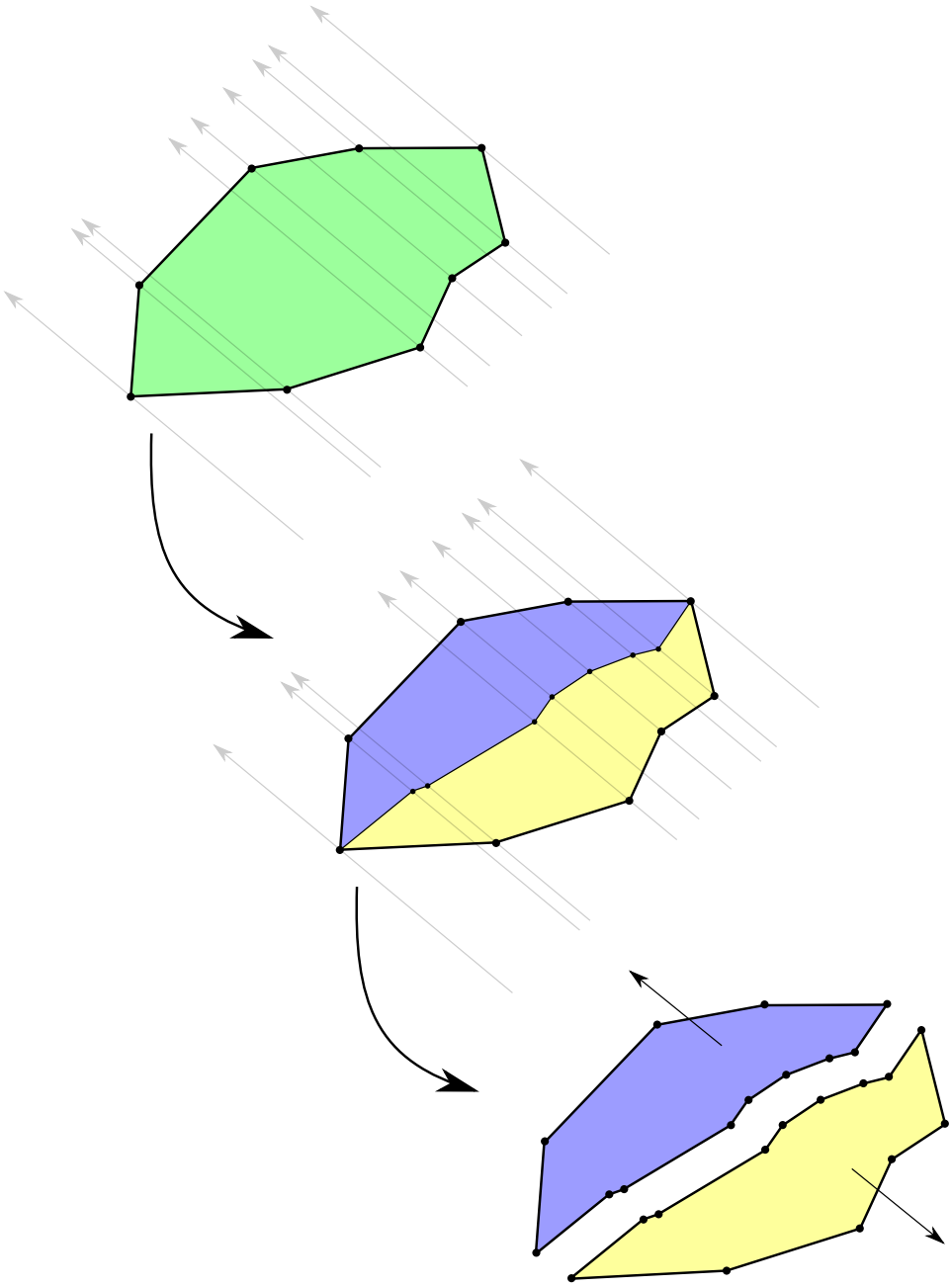


Figure 10.2: Demonstration of projective segmentation on a well-behaved simple polygon. The casting direction is indicated by gray arrows and the fractional width is 1/2. Figure 4.8 shows recursive applied projective segmentation.

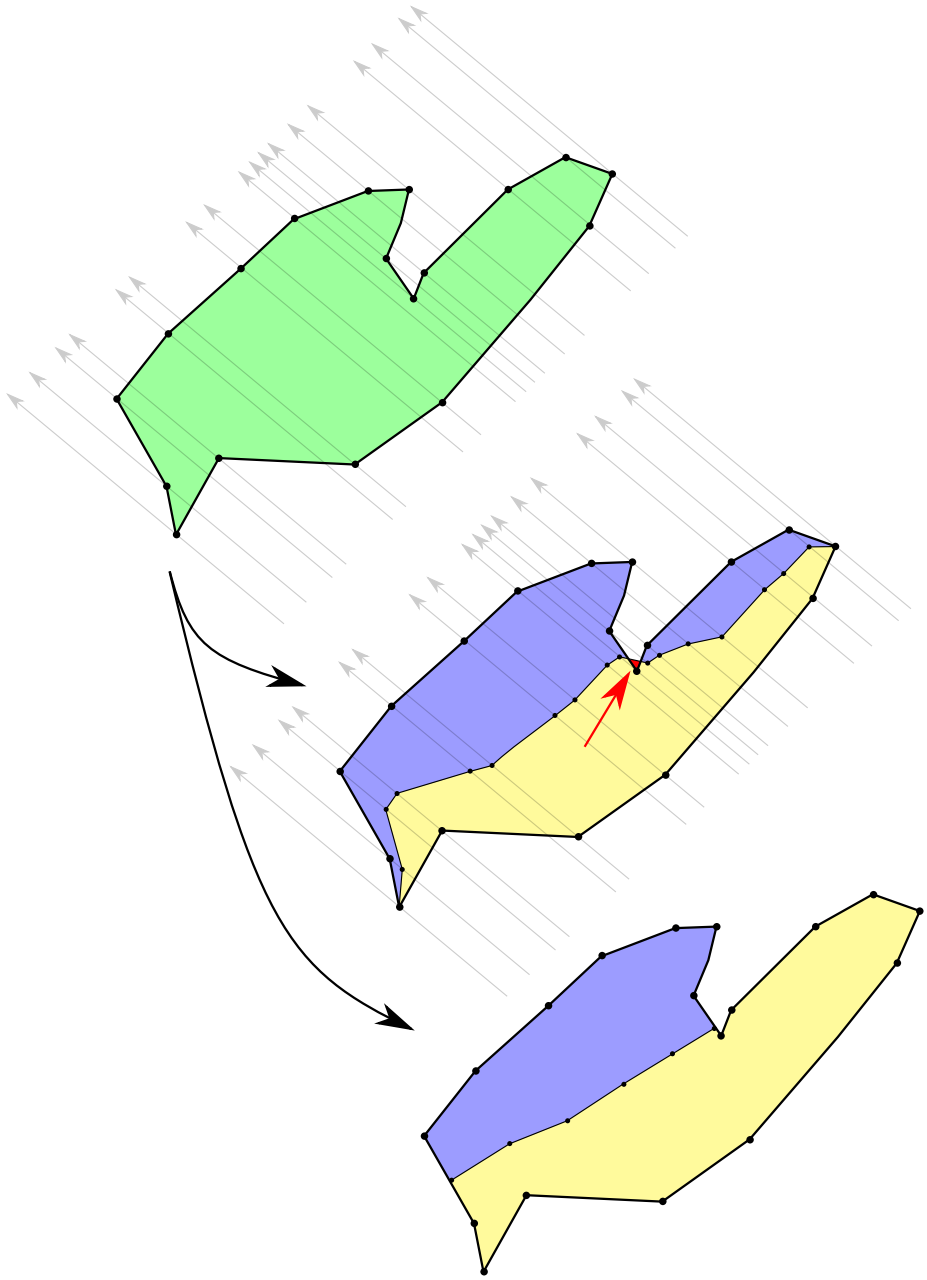


Figure 10.3: Unintuitive 'jumps' and the failure of projective segmentation on a simple polygon (indicated with an arrow). The top figure is projectively segmented into the middle figure. Note the cleave line passes outside the original polygon. The bottom figure represents a cleave that is more intuitive, but not attainable using projective segmentation.

10.6 Iterative Segmentation

If one knows the location and orientation of a desired cleaving plane beforehand (e.g., centroid and an axes-aligned plane intersecting the ROI centroid), then planar segmentation can be used directly. However, often the goal is to sub-segment based on volumetric criteria so that a specific fraction of resultant sub-segment volume is above (or below) a given plane. Projective segmentation is able to incorporate this criteria as the fractional width parameter and tautologically find sub-segments of a given fractional volume (up to the area/volume conservation discrepancy), but planar segmentation cannot. In this case the offset of the plane is unknown (but the orientation is assumed to be known).

Determination of the planar offset can, in principle, be solved analytically, however the solution must account for all ROI vertices and their connectivity, and is therefore cumbersome to specify in closed form. An alternative approach is to create an objective function based on the volumetric criteria and optimize the planar offset. In practice, iterative bisection converges quickly, reliably, is guaranteed to make forward progress at each iteration, and can directly use a volume estimate objective function.

The algorithm proceeds by specifying the plane orientation (via a unit vector perpendicular to the final plane), a desired volume constraint (e.g., the sub-segment on the positive side of the final plane should contain 23.46% of the whole ROI(s) volume), and a suitable tolerance (e.g., 0.05%). First, two planes with the given orientation that bound the whole ROI(s) are found by scanning for extremal vertices. (An additional margin can be added to improve convergence in some cases.) The volume of the (whole) ROI(s) is computed. Then, a loop is entered. First, the plane intersecting the midpoint between the bounding planes is computed. The ROI(s) is segmented along it (planar segmentation). The volume of one sub-segment is computed and compared to the desired constraint. If it is within tolerance, segmentation is considered sufficient and iteration stops. Otherwise, one of the bounding planes is replaced with the midpoint plane depending on whether the segmented volume is above or below the criteria. The loop is

iterated until the stopping criteria is met.

Bi-section is useful because it can be applied to any planar segmentation scenario; it does not need to be adapted in any way except to find useful initial bounds which is a minor, straightforward problem. Bi-section as applied in this context can be thought of as a branch-and-bound algorithm [421]; the number of candidate planes are fixed spatially by the mid-point and harmonics of the original bounding planes and form an infinite tree of potential solutions. The replacement of a bound by the midpoint at each step guarantees forward progress and also prunes the tree of unsuitable candidates which are then never visited. These well-known characteristics have caused branch-and-bound methods to be widely applied for challenging computational problems for which no polynomial-time algorithms are known [422, 423, 424]. No prior record of application to the problem of ROI segmentation was found in the literature.

10.7 Conclusions

`DICOMautomaton` implements a novel 'vector' approach to ROI segmentation. Iterated bi-section of planar segmentation provides an efficient and robust way of partitioning ROI into sub-segments with specific volumetric criteria; only a cleaving plane orientation and tolerances must be specified.

Chapter 11

Segmentation Methodology[1]

11.1 Introduction

Heterogeneous functional dose response for OARs is becoming increasingly relevant for clinical radiotherapy planning. In 2005, Konings et al. [348] found evidence of region-dependent volume effects in rat parotid. Years later in 2010, as part of the encompassing *Quantitative Analysis of Normal Tissue Effects in the Clinic* (QUANTEC) organ-focused reviews for clinical guidelines, Deasy et al. [236] concluded that better predictive models were needed to model xerostomia risk. One factor recommended for investigation was whether regions within the parotid could be located that exhibited variable dose sensitivity, increased or decreased functional burden, or otherwise controlled function preservation to a higher degree than surrounding tissues. Other articles in the same report provided similar recommendations for other organs [425, 426, 427]. In response, organ ROIs are increasingly being segmented or handled heterogeneously to model dose response to various aspects within the organ. Reports of trials underway are emerging [357, 358, 428].

Methods for complex contour segmentation, including planar segmentation, have been demonstrated in the literature [7] and were described in

[1] The contents of this chapter and appendix A were submitted to *Physics in Medicine and Biology* under the title '*Prefer Nested Segmentation to Compound Segmentation*' on the 27th of October, 2016. The list of authors was Haley Clark, Stefan Reinsberg, Vitali Moiseenko, Jonn Wu, and Steven Thomas.

chapter 10. A recent paper by van Luijk et al. [357] made use of a planar segmentation method in which fractional volumes were used to implement a compounded Boolean sub-segment selection mechanism. Here we show that, perhaps unintuitively, such a scheme will result in sub-segments of differing volume depending on the shape of the ROI and selection location within it. Inconsistent segmentation volumes can be problematic for investigation of sub-organ effects because sub-segments will represent inconsistent portions of the whole ROI. Performing sensitivity analysis, model fitting, or tests of associativity (e.g., correlation) will result in bolstered or undermined sub-segment importance, model parameters, or associativity, which must be corrected.

We propose an improvement, which we call *nested segmentation*, that is "fair" in the sense that it will produce equal-volume sub-segments uniformly throughout the ROI when the cleaving method is free of bias. Furthermore, it is robust – if the cleaving method *is* biased, as-fair-as-possible sub-segments are produced. It is also faster than compound segmentation, requiring equivalent or less geometrical processing. An implementation based on segmentation of planar contours is tested using clinical data from 510 head-and-neck cancer patients.

We also present two methods that can be used in conjunction with segmentation that help ensure an equal number of grid voxels (e.g., radiotherapy dose matrix voxels) are contained within the boundaries of each sub-segment: oblique cleaving planes and grid supersampling. We show that both methods ameliorate issues arising from collinearity of dose grid and ROI boundaries.

11.2 Materials and Methods

11.2.1 Segmentation

Segmentation[2] refers to the process in which part of a volume delineated by closed contour lines (i.e., a ROI) is partitioned into connected pieces ("sub-segments") and one or more are retained ("selected"). We refer specifically to

[2] Alternatively *contour sub-segmentation* or just *sub-segmentation*, to differentiate it from *image* segmentation.

segmentation, but the process is equivalent to *volume truncation* for polyhedra and generic *division* or *partitioning* of areas, volumes, and hypervolumes (e.g., geometric primitives, such as triangles, spheres, and cubes). Sub-segment selection can be accomplished in a variety of ways, but in this work we focus on the method described by van Luijk et al. [357], which we refer to as *compound* segmentation. Our improved method, nested segmentation, is believed to be more robust toward the 'fair distribution' problems of ensuring that selected sub-segments have equivalent volume and contain an equivalent number of entities (e.g., dosimetric grid voxels) regardless of the ROI shape and selection location (e.g., periphery vs. centre). Salient differences are described in the following subsections.

11.2.2 Compound Segmentation

Compound segmentation is a planar segmentation technique that makes use of (infinite) cleaving planes. The cleaved sub-segment faces are flat, and thus when the ROI is convex all cleaved surfaces remain convex. Compound segmentation proceeds by specification of cleaving plane orientations and volume percentiles (i.e., fractional volumes). Each fractional volume ($f \in [0, 1]$) unambiguously specifies a cleaving plane which contains f on one side of the plane and $1 - f$ on the other[3]. Each plane requires a single unit vector or two free parameters to orient the plane. In [357] six planes are located and used to select a sub-segment interior to the boundary of a parotid ROI. There are three sets of parallel planes; each set is orthogonal to the others (see figure 11.1). Use of one less plane would permit selection of an arbitrary sub-segment with a single portion of the ROI surface[4], use of two less planes would permit either one or two disjoint ROI surface portions, etc..

In compound segmentation, all cleaving planes are derived using percentiles or fractional volumes that refer to the whole ROI. Only after all planes are located is segmentation performed by application of cleaving planes to the ROI volume, and only the interior is selected.

[3]Both $f = 0$ and $f = 1$ are ambiguous because they are not unique. The ambiguity is not relevant for segmentation.

[4]In the case of convex ROIs.

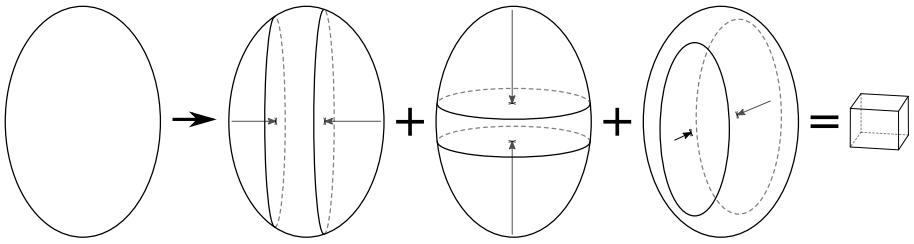


Figure 11.1: Demonstration of compound segmentation with three parallel pairs of mutually orthogonal planes (six planes in total).

11.2.3 Nested Segmentation

We propose an improved method in which sub-segments are selected as the interior region between two parallel planes, as with compound segmentation, however cleaves are performed eagerly, before the next pair of planes can be located. The location of cleaving planes are thus derived from the volume of *remaining* sub-segments, not the original ROI (see fig. 11.2). As each individual stage of segmentation achieves a fair divvy of the remaining volume, sub-segments are expected to always contain an equivalent portion of the ROI volume when the partitioning method is fair.

11.2.4 ROI Segmentation Comparison

We compare segmentation methods on a data set of 510 parotid gland ROIs from 510 head-and-neck cancer patients (one per patient, to avoid any potential bias from shape correlation between left and right parotid). We perform segmentation to generate 3, 18, and 96 sub-segments from each ROI, chosen so that sub-segments are not excessively elongated along any one linear dimension but the discrete axial nature is accommodated. 18^{ths} and 96^{ths} segmentation used 6 cleaving planes (i.e., three mutually orthogonal sets of parallel planes) for each sub-segment whereas segmentation into thirds used a single pair of planes parallel to the ROI contours. Because ROIs are defined in terms of equidistant, parallel, planar contours with no gaps, contour area is used as a surrogate for volume.

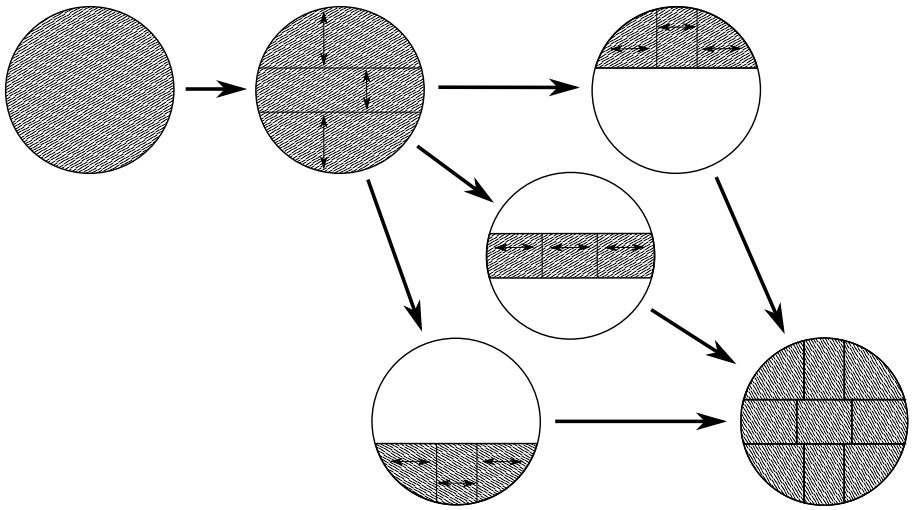


Figure 11.2: Demonstration of nested segmentation on a circle with $f = 1/3$. All sub-segments have area $\pi r^2/9$.

Each parotid in this data set includes a radiotherapy treatment planning dosimetric grid which is used to derive dose-volume statistics of interest. We compute the number of voxels lying within each sub-segment and compare distributions for each sub-segment. Cardinal axes-aligned cleaving planes are often desired for ease of specification or bounding of anatomical regions (e.g., posterior region, lateral-caudal region, etc.). However, raster grids are commonly aligned with the cardinal directions, and are in this data set, which makes perfectly fair partitioning of voxels impossible (i.e., due to collinearity; a row locally aligned with the contour boundary is either within or outside of the sub-segment, but the row may contain many voxels). We employ two techniques which can help to more fairly partition sub-segments: raster grid supersampling and oblique cleaving planes. Supersampling used fine ($15\times$) cubic interpolation so that each voxel was effectively interpolated into 225 voxels. A cyclic rotation of 22.5° between cardinal axes was used to orient oblique planes.

Contours are operated on directly rather than rasterizing them onto a volumetric grid. Bisection is used to locate planes corresponding to the requisite f. The bisection method used for $\mathbb{R}^3$ ROI segmentation had a

stopping tolerance set to 1% for all clinical data segmentation, but contours lying in planes parallel to the cleaving plane were treated atomically and were thus indivisible. Parotids with few contours therefore could not achieve 1% tolerance. Voxels for each dose matrix were of fixed volume, so summary statistics about the distribution of sub-segment voxel counts estimate sub-segment volume.

Segmentation into thirds employed only two planes parallel to contours. Nested and compound segmentation should produce identical results in this case because only a single pair of cleaves are performed. However, it is challenging because contours are not divided in this case. We include the comparison to demonstrate that bisection is produces sufficiently fair sub-segments. All ROI and dose manipulations were performed using `DICOM-automaton` [7, 411].

11.2.5 Statistics

Distributions of volumes and voxels counts for sub-segments were compared using a non-parametric Kolmogorov-Smirnov test. The null hypothesis is that one of the distributions is drawn from the same parent distribution as the other, so that the distributions are statistically identical. Individual sub-segment volumes are compared with other sub-segments in the same ROI using voxel counts to determine the spread due to the cumulative effects of the segmentation method and raster grid voxel alignment. The Quartile Coefficient of Dispersion (QCD) provides a normalized measure of the dispersion of sub-segment areas for each patient [429]. Median QCD and median-normalized ranges are reported to characterize the population. A standard statistical significance threshold (α) of 0.05 was used.

11.3 Results

11.3.1 Analytic Comparison

The $\mathbb{R}^2$ analog to ROI segmentation is individual planar contour segmentation. We segment a circle of radius r into nine sub-segments using compound

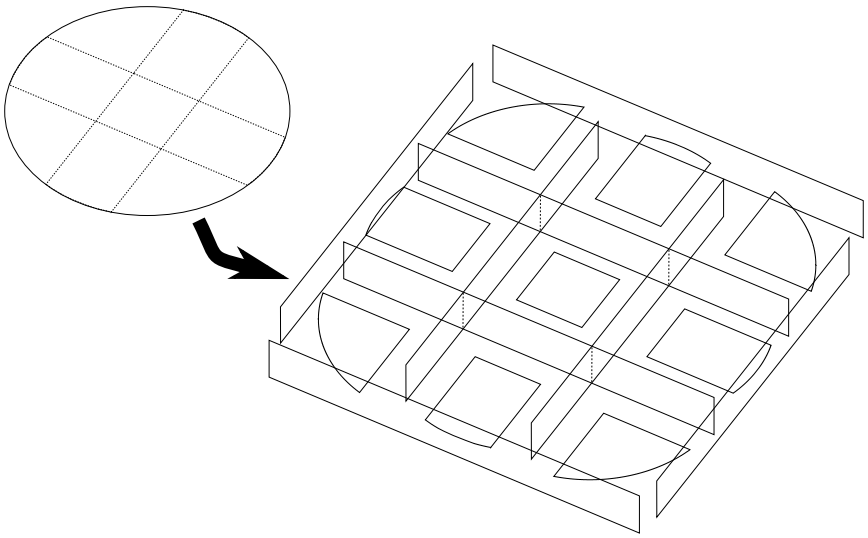


Figure 11.3: Partitioning of a circle into nine sub-segments using compounded segmentation (exploded view). Each sub-segment is bounded by two parallel pairs of mutually orthogonal planes.

segmentation (see fig. 11.3). The fractional area on the small side of each cleaving plane is 1/3. Cleaving plane orientations are fixed, but the offsets from the origin are unknown and are derived analytically or located through, e.g., bisection. Using elementary methods, it can be shown that the fractional area enclosed by a plane offset from the origin (n.b. a secant line) and a parallel plane intersecting the origin, in terms of their separation ($h \in [0, r]$; i.e., the *apothem*; cf. [430]) is

$$f = \frac{2}{\pi}\left(\arcsin\frac{h}{r} + \frac{h^2}{r^2}\cot\arcsin\frac{h}{r}\right). \tag{11.1}$$

Inversion is used to determine h. When $f = 1/3$, $h \approx 0.264932r$. Derivation of the nine sub-segment areas is then straightforward (see fig. 11.4). Results are summarized in table 11.1. The smallest, as a ratio of the 'fairly distributed' area ($\pi r^2/9$) is the centre sub-segment at ≈ 0.8043; the centre-adjacent sub-segments are the largest at ≈ 1.0978.

Nested segmentation, on the other hand, generated sub-segments with

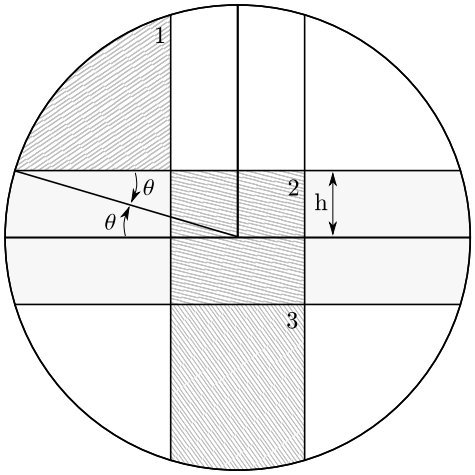


Figure 11.4: Calculation of sub-segment areas in terms of the area of a wedge, right triangle, and square defined by $f = 1/3$, $h \approx 0.264932r$, and r. Three distinct types of sub-segments are shown: (1) "corner," (2) "centre," and (3) "centre-adjacent."

	Ratio of Fair	
Sub-segment	(general f)	($f = 1/3$)
centre	$\frac{36}{\pi}\left(\frac{h}{r}\right)^2$	$0.804306\times$
centre-adjacent	$\frac{9}{2}f - \frac{18}{\pi}\left(\frac{h}{r}\right)^2$	$1.097847\times$
corner	$\frac{9}{4}(1-2f) + \frac{9}{\pi}\left(\frac{h}{r}\right)^2$	$0.951077\times$

Table 11.1: Ratios of the fair fractional area for compound segmentation sub-segments in terms of the apothem (h) and fractional area (f). All ratios are fractions of the fairly distributed area ($\pi r^2/9$) in which each sub-segment has an equivalent area. Centre sub-segments have four planar edges, centre-adjacent have three, and corner sub-segments have two.

equal area (see fig. 11.5). If all partitions can be made fairly, so that a cleaving plane that achieves the desired fractional areas is located exactly, then each sub-segment area is tautologically known as the product of requested fractional areas. For example, each final sub-segment in figure 11.2 has an area 1/3 of 1/3 of the total. The first cleave is identical to compound segmentation

and so the apothem is given by eq. (11.1). Because nested segmentation is a greedy algorithm and the first cleave does not take into account later cleaves, sub-segments are in general asymmetric. The two asymmetries possible for segmentation into nine sub-segments (n.b. with fixed cleave plane orientations) are shown in fig. 11.5.

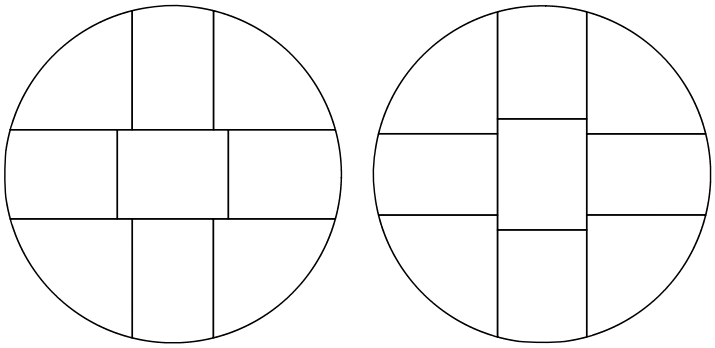


Figure 11.5: Nested segmentation of a circle into nine sub-segments each with area $\pi r^2/9$. The orientation of the first cleave can be chosen two ways. Both are shown. The cleaving order is important in nested segmentation, but not for compound segmentation.

Moving to $\mathbb{R}^3$, compound segmentation applied to a sphere partitioned into $3 \times 3 \times 3 = 27$ sub-segments yielded a centre sub-segment volume ≈ 0.596 that of the fair volume. Centre-adjacent-adjacent sub-segments had areas ≈ 1.105 that of the fair volume. Nested segmentation again produced fair volumes that were tautologically, in this case, $1/27^{\text{th}}$ of the whole. A sphere constructed of discrete stacks of contours sharing a planar orientation, which is common for medical image ROIs, approached both compound and nested segmentation results asymptotically as the contour thickness shrunk.

11.3.2 Segmentation into Thirds

Using compound segmentation, whole ROI were segmented into three sub-segments (f_{axial} spanned $\left[0, \frac{1}{3}\right]$, $\left[\frac{1}{3}, \frac{2}{3}\right]$, and $\left[\frac{2}{3}, 1\right]$). Bisection was employed and cleaving planes were held parallel to contours. The median number of voxels in each sub-segment spanned $587.5 - 605.0$. The distribution of voxel

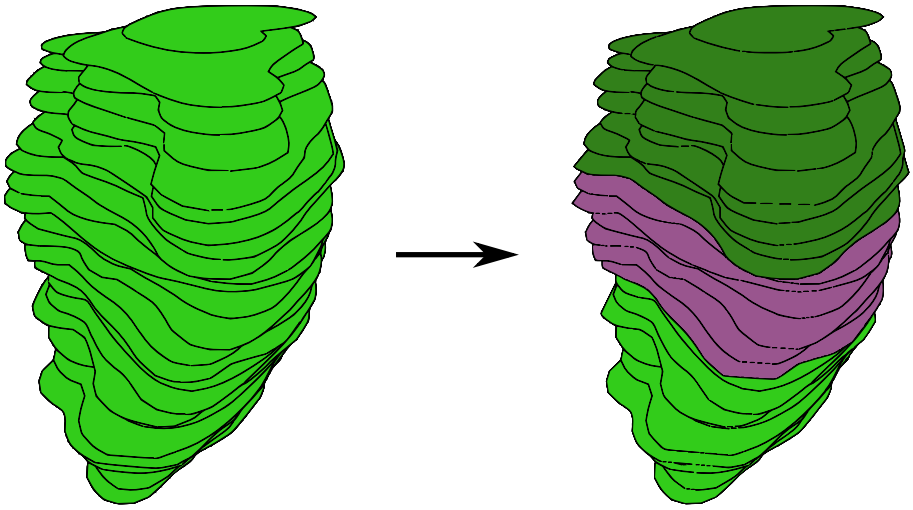


Figure 11.6: Depiction of the way in which the parotid gland ROI volume was segmented to achieve sub-segments with volume 1/3 that of the whole parotid. Nested and compound segmentation produce identical results in this case.

counts in cranial, middle, and caudal sub-segments were compared with a Kolmogorov-Smirnov test. Each unique comparison in {left, right} parotid $\otimes$ {cranial, middle, caudal} sub-segments was performed, yielding ten tests. In all cases the two-sided $p > 0.20$. These tests indicate the bisection approach results in appropriately partitioned sub-segments that contain 1/3 of the original parotid volume without systematic bias detectable at the $\alpha = 0.05$ level. Results were identical for nested segmentation.

11.3.3 Segmentation into 18$^{\text{ths}}$

Figure 11.7 shows nested and compound segmentation of whole parotid into 18 sub-segments. Sub-segments are composed of axially-adjacent slices coloured uniformly[5]. Both f_{axial} and f_{sagittal} spanned $\left[0, \frac{1}{3}\right]$, $\left[\frac{1}{3}, \frac{2}{3}\right]$, and $\left[\frac{2}{3}, 1\right]$; f_{coronal} spanned $\left[0, \frac{1}{2}\right]$ and $\left[\frac{1}{2}, 1\right]$. The cleaving order was axial $\rightarrow$ coronal $\rightarrow$ sagittal. As can be seen in figure 11.7, nested and compound method sub-segment

[5]Colours were chosen for maximum contrast using a modification of the palette described in [431].

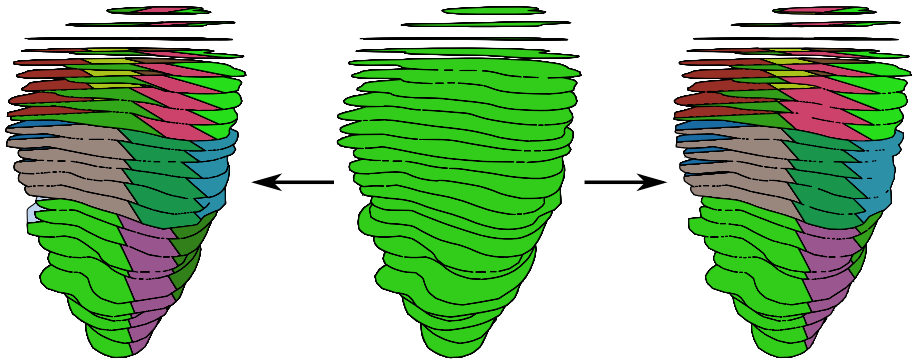


Figure 11.7: Depiction of nested (left) and compound (right) segmentation of whole parotid (centre) into 18 sub-segments.

locations differ only slightly. However, it is apparent that sub-segments in the compound method do not all have equivalent volume.

Using compound segmentation without supersampling or oblique cleaving planes, sub-segment voxel counts had a mean of 100.0 voxels within each sub-segment (std. dev. = 64.0; std. dev. of the mean = 0.7; median = 92.0). The median number of voxels in each sub-segment spanned $38 - 152$. Only 57.6% of sub-segment voxel counts had an absolute percent difference of less than 50% of the mean. Direct comparison of the voxel count distributions within sub-segments was performed using Kolmogorov-Smirnov tests. Unique comparison of all 18 sub-segments required 153 tests – in 124 cases (81%) the null hypothesis failed to be rejected and distributions were found to differ significantly (i.e., $p < 0.05$ in 124 cases). Skewness of the combined voxel count distribution was 0.991 using the ratio of moments technique, which indicates a strong positive skew. However, the mean voxel count in each *type* of sub-segment were more symmetrically distributed with a skewness of 0.075 and a std. dev. = 28.0. No significant correlation was detected between the average sub-segment mean voxel count and position relative to the parotid centre (e.g., with relativity denoted by -1, 0, or $+1$ in the cardinal directions).

Using nested segmentation without supersampling or oblique cleaving planes, sub-segment voxel counts again had a mean of 100.0 voxels within

each sub-segment (std. dev. = 49.8; std. dev. of the mean = 0.6; median = 97.0). However the median number of voxels in each sub-segment spanned $92.5 - 101$. 70.2% of sub-segment voxel counts had an absolute percent difference of less than 50% of the mean. Direct comparison of the voxel count distributions within sub-segments using Kolmogorov-Smirnov tests showed that the null hypothesis failed to be rejected in only 2 of 153 (1.3%) cases (i.e., $p < 0.05$ in 2 cases).

11.3.4 Segmentation into 96^{ths}

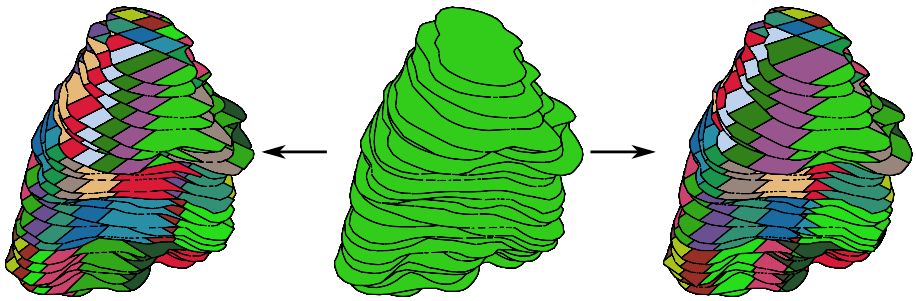


Figure 11.8: Depiction of nested (left) and compound (right) segmentation of whole parotid (centre) into 96 sub-segments.

For segmentation into 96^{ths}, both f_{axial} and f_{coronal} spanned $\left[0, \frac{1}{4}\right]$, $\left[\frac{1}{4}, \frac{1}{2}\right]$, $\left[\frac{1}{2}, \frac{3}{4}\right]$, and $\left[\frac{3}{4}, 1\right]$ whereas f_{sagittal} spanned $\left[0, \frac{1}{6}\right]$, $\left[\frac{1}{6}, \frac{1}{3}\right]$, $\left[\frac{1}{3}, \frac{1}{2}\right]$, $\left[\frac{1}{2}, \frac{2}{3}\right]$, $\left[\frac{2}{3}, \frac{5}{6}\right]$, and $\left[\frac{5}{6}, 1\right]$. The cleaving order was axial $\rightarrow$ coronal $\rightarrow$ sagittal. As can be seen in figure 11.8, nested and compound segmentation again produce similar-shaped sub-segments in roughly similar locations. However, compound segmentation produces sub-segments with substantially different volumes, such as those with vanishingly small volume (in the centre-bottom of fig. 11.8; right side). The comparable nested segmentation sub-segments, on the other hand, are larger and have the same apparent volume as all other sub-segments (left side of fig. 11.8). Nested method sub-segment median QCD were less disperse than the compound method (0.097 vs. 0.37). Oblique planes reduced dispersion nearly 25$\times$ for nested method sub-segments ($0.097 \rightarrow 0.0041$). Conversely, they *increased* compound method dispersion ($0.37 \rightarrow 0.46$).

	Compound	Nested			
	Oblique + Supersampling	Unmodified	Oblique	Supersampling	Oblique + Supersampling
Median range	$91.5 - 13829$	$20 - 30$	$18 - 20$	$4027.5 - 4200$	$4212 - 4287$
Median	5558	24	19	4103	4246
Range/Median	2.47	0.417	0.105	0.042	0.018
Sig. K-S tests	2435 (53.3%)	258 (5.7%)	1 (0.02%)	0 (0%)	0 (0%)
QCD	0.46	0.097	0.0041	0.097	0.0041
Runtime	$143\,ms$	$36\,ms$	$29\,ms$	$135\,ms$	$131\,ms$

Table 11.2: Comparison of median voxel counts, quartile coefficients of dispersion (QCD), and runtime for compound and nested segmentation. Sig. K-S tests refers to the number of statistically significant Kolmogorov-Smirnov tests (out of 4560; $\alpha = 0.05$). Runtime is per (individual) sub-segment and was measured on an Intel® Xeon® X5550 CPU. The use of oblique cleaving planes and fine supersampling reduced sub-segment median voxel range relative to the median.

A comparison of voxel counts and runtime for compound and nested segmentation is summarized in table 11.2. Using compound segmentation without supersampling or oblique cleaving planes led to unusable data; for each sub-segment, at least one patient had a vanishingly small sub-segment encompassing zero voxels. Using fine supersampling and oblique planes, sub-segment voxel counts had a mean of 6244.5 voxels within each sub-segment (std. dev. = 4417.6.0; std. dev. of the mean = 96.1; median = 5558.0). The median number of voxels in each sub-segment spanned $91.5 - 13829$, encompassing two orders of magnitude. Direct comparison of the voxel count distributions within sub-segments via Kolmogorov-Smirnov tests showed the null hypothesis failed to be rejected and distributions were found to differ significantly in 2435 of 4560 (53.3%) unique test cases (i.e., $p < 0.05$ in 2435 cases).

Nested segmentation was markedly different. Using nested segmentation *without* supersampling or oblique planes, sub-segment voxel counts had a mean of 27.2 voxels within each sub-segment (std. dev. = 12.3; std. dev. of the mean = 0.15; median = 24.0). The median number of voxels in each sub-segment spanned $20 - 30$. Direct comparison of the voxel count distributions within sub-segments via Kolmogorov-Smirnov tests showed the null hypothesis failed to be rejected and distributions were found to differ significantly in 258 of 4560 (5.7%) unique test cases (i.e., $p < 0.05$ in 258 cases). Applying the oblique planes method yielded a mean sub-segment voxel count of 19.4 voxels within each sub-segment (std. dev. = 9.0; std. dev. of the mean = 0.04; median = 19.0). The median number of voxels in each sub-segment spanned $18 - 20$. Direct comparison of the voxel count distributions within sub-segments yielded significance in a single case out of 4560 (0.02%; i.e., $p < 0.05$ for the Kolmogorov-Smirnov test in one case). Applying supersampling with axis-aligned planes yielded a mean sub-segment voxel count of 4199.3 voxels within each sub-segment (std. dev. = 2139.3; std. dev. of the mean = 9.7; median = 4103.0). The median number of voxels in each sub-segment spanned $4027.5 - 4200$. No voxel count distributions were significantly distinct according to the Kolmogorov-Smirnov test (i.e., $p < 0.05$ in *zero* of 4560 tests).

Using both oblique planes and supersampling improved fairness of nested segmentation even more, though either oblique planes or supersampling alone were sufficient for most purposes. The mean sub-segment voxel count was 4270.2 voxels within each sub-segment (std. dev. = 2004.8; std. dev. of the mean = 9.1; median = 4246.0). The median number of voxels in each sub-segment spanned $4212 - 4287$. Direct comparison of the voxel count distributions again found no significantly distinct distributions (i.e., $p < 0.05$ in *zero* of 4560 tests).

11.4 Discussion

Planar segmentation can be accomplished using a variety of existing tools, e.g., via Boolean structure combination [432, 433], conversion of ROIs to polygon surface meshes and computing the intersection [414, 415] or via tessellation [434], or directly on ROI contours via bisection [7]. Whatever the method, sub-segments are effectively specified by the fractional volume between mutually orthogonal pairs of cleaving planes. It may then seem intuitive that sub-segments with the same fractional volume between bounding planes, but at different positions in the ROI, would have the same volume. This intuition is valid for nested segmentation, but not for compound segmentation. Compound segmentation only generates fair sub-segments when the ROI is rectangular and faces are aligned with the cleaving planes, and thus may lead to erroneous conclusions if used for sub-segment comparison. A number of articles investigating the link between patient outcomes and radiotherapy dose to parotid sub-volumes have recently emerged [324, 358, 428] and use of compound segmentation has been reported in the literature [357]. The aim of this study was to demonstrate that nested segmentation is fairer than compound segmentation, and should be preferred for analyses involving sub-segment comparison.

By analytically solving $\mathbb{R}^2$ and $\mathbb{R}^3$ analogues, we showed that compound method sub-segments have intrinsically non-uniform area/volume. In $\mathbb{R}^2$, compound method centre sub-segment area differed from that of adjacent sub-segments by nearly a *third* of the fair area. The problem grew worse

in $\mathbb{R}^3$ with the difference assuming more than *half* the fair volume. Nested method sub-segments were fair in both cases.

The successful segmentation of clinical ROIs into thirds indicates bisection is appropriate for locating cleaving planes despite being unable to fairly partition due to discrete nature of contours along the axial direction. Compound segmentation into 18^{ths} was not fair. Distribution skewness and distinctness tests imply that the parotid was not fairly partitioned into sub-segments of equivalent volume. At the same time, the lack of correlation between sub-segment mean voxel count and relative position indicates the bisection approach is not systematically biasing results and that sub-segment volumes appear to be comparable *on average*. Nested segmentation, in comparison, was fair. Distribution distinctness test results were substantially improved compared with compound segmentation (2 vs. 124 of 153 tests found distinct distributions). The low number of distinct distributions (1.3%) was comparable with the bisection tolerance (1%) and therefore represents an acceptable deviation. Performance on the Kolmogorov-Smirnov test is notable because the transverse cleave generally can not achieve fair cleaving. The transverse cleave was performed first, and it is apparent that subsequent cleaves are fairer than those of compound segmentation.

The distinction between compound and nested segmentation was embiggened by segmentation into 96^{ths}. Some peripheral compound method sub-segments with vanishingly small volumes – even when oblique planes and intensive supersampling were employed. Nested method sub-segments were not quite fair when oblique planes and supersampling were abstained from (5.7% of Kolmogorov-Smirnov tests were significant), but this was corrected when oblique planes, supersampling, or both were employed (0.02% or less in all cases). The normalized range of voxels contained within a sub-segment dropped when using oblique planes or supersampling, indicating sub-segment volumes became fairer when either were employed. Compared to compound segmentation, nested segmentation produced normalized ranges that were two orders of magnitude smaller. Additionally, QCD differed by 1-2 orders of magnitude depending whether oblique planes were used, suggesting nested method sub-segments were substantially less disperse, and thus more uniform,

than compound method sub-segments. These observation support the claim that nested segmentation is resilient to partitioning errors. When a fair cleave could not be located, i.e., due to discrete nature of contours along the axial direction, child sub-segments were made as fairly as possible (i.e., sub-segments equally shared the remaining volume with sibling sub-segments). The increase of dispersion noted in compound method sub-segments when oblique planes were used supports the claim that compound segmentation intrinsically can not fairly partition ROIs.

Nested segmentation was not only fairer than compound segmentation, but it was also faster. Sub-segments have planar edges/faces that can be described with few vertices. In nested segmentation, recursive segmentation need only process a simplified geometry for each planar edge/face. The full ROI is processed once; afterwards, each additional segmentation is continually reduced by the increasing number of planar edges. Compound segmentation, however, must continually re-process the full ROI. The exact speed-up depends on ROI geometry, contour sampling density, and nesting depth.

One downside of nested segmentation is that the shape of sub-segments depends on the order of cleaves, resulting in shape asymmetries. For example, when partitioning a circle into 1/6 using three orthogonal sets of parallel cleaving planes, both compounded and nested segmentation perform the same first cleave ($f = 1/3$), which results in a plane with distance $h \approx 0.264932r$ from the centre of the circle. The compound method finds the second cleave in the orthogonal direction to have the same distance from the centre, which results in lower-than-fair sub-segment area, whereas the nested segmentation method finds $h \approx 0.329392r$, which makes the sub-segment area fair ($= \pi r^2/9$) but is asymmetric. A perfectly symmetrical method may be possible and would find $h = r\sqrt{\pi/36} \approx 0.295408972r$ in both directions, but would require advance knowledge of all intended fractional volumes and cleaving planes and would most likely require an iterated relaxation step. A perfectly symmetrical segmentation method may be possible, but would most likely require iteration or back-tracking and re-processing whole ROIs for each sub-segment. In contrast, nested segmentation requires neither back-tracking nor re-processing geometry. Nested segmentation is directly applicable to

organs where anatomical structure is ignorable or *a priori* unknown. It can also be employed within larger anatomical groupings, such as within lobes or cavities (e.g., liver, lung), and can make use of oriented cleaving planes or shuffled cleaving orders that align with local anatomy (e.g., muscle tissues, vessels, ducts). The use of planar segmentation combined with (iterated) bisection is a flexible paradigm that enables the use of individual $\mathbb{R}^2$ contours, raster grids, disconnected collections of contours, contours with holes, and volumetric surface manifolds, and would therefore be suitable addition to software packages that can potentially operate on any such primitives (e.g., [7, 433]).

It is worthwhile to compare an alternative routine that would use relaxation to adjust sub-segment locations based on a penalty function, heuristic, or clustering. The method of nested segmentation is fast and requires no backtracking or recomputation. A technique based on clustering or optimization would likely require iterated relaxation. An improved method would be aware of the entities within the sub-segments (e.g., voxel count) but as shown it may be impossible to make perfectly fair partitions even if there is specific knowledge of each entity. Nested segmentation is therefore not substantially worse than a more advanced algorithm, but is extremely easy to implement and verify.

Oblique cleaving planes addressed the issue of ROI segment and voxel grid collinearity, but can result in awkward plane orientations in some cases. There is an optimal cleave plane orientation that can be determined exactly when sub-segment extents are known. The optimal angle is found for some special cases in $\mathbb{R}^2$ in a appendix A. This orientation maximizes the minimum spacing between voxel distances to the plane, ensuring small changes in the plane position results in the smallest possible number of voxels crossing the plane at one time (e.g., minimizing spatial resonances). Unfortunately, even estimation is difficult and costly [435, 436] so throughout this work a cyclic rotation of 22.5° between cardinal axes defined by the Cartesian dose grid was assumed. Supersampling is also useful for improving sub-segment fairness, though it can not itself help the collinearity issue if planes are axes-aligned. However, when oblique planes and supersampling are combined,

supersampling will reduce the amount of obliquity needed, which can assist in adapting to underlying anatomy. It will also result in sufficiently fair sub-segments if supersampling can be performed to an arbitrary level, though it is computationally difficult and questionable to supersample too finely. Oblique planes were more computationally efficient than supersampling, but application of either method independently for nested segmentation into 96^{ths} resulted in small median voxel ranges and acceptably indistinct distributions (i.e., $> 99.9\%$ with $p > 0.05$).

11.4.1 Conclusions

Nested segmentation was found to be superior to compound segmentation when sub-segment volume consistency is needed.

Chapter 12

Parametric Approach to Regional Effect Assessment[1]

12.1 Introduction

Current clinical guidelines recommend treating the parotid gland as a parallel organ and advocate use of whole parotid mean dose thresholds for sparing [236]. However, mounting evidence suggests there may be dosimetrically critical sub-structures within the parotid that strongly influence production of saliva after radiotherapy [324, 357, 358].

Volume effects in parotid have a storied history, and many attempts have been made to locate critical regions and elucidate the mechanism(s) of dysfunction. In 2005, prompted by the spread of IMRT and the corresponding need for assessment of independent functional unit distribution within the parotid, Konings *et al.* reported observing region-dependent late volume effects in rat parotid [348]. Dose delivered to cranial and caudal parotid halves resulted in different outcomes; salivary dysfunction was more pronounced after irradiation of the cranial aspect. Additionally, cranial aspect irradiation

[1]The contents of this chapter have been submitted to *Acta Oncologica* under the title '*Caudal Aspects of the Parotid Gland are Most Important for Radiation-Induced Salivary Dysfunction*' on the 15$^{\text{th}}$ of November, 2016. The list of authors was Haley Clark, Steven Thomas, Jonn Wu, Allan Hovan, Carrie-Lynne Swift, and Stefan Reinsberg.

caused secondary damage to appear in the non-irradiated caudal aspect. Konings *et al.* surmised that the effect was caused by injury to major excretory ducts, blood supply, and innervation in the region between ventral and dorsal lobes. Dijkema *et al.* reported in 2008 that mean dose models anomalously failed to describe human parotid function loss when comparing conventional radiotherapy to IMRT [355]. In particular, dose-response 'left-shifted' compared to conventional radiotherapy, meaning lower doses were required to achieve the same loss of function. They also hypothesized that the pervasive low dose spread by IMRT throughout the parotid might directly cause dysfunction via acinar cell plasma membrane damage. In 2009, van Luijk *et al.* noticed that earlier studies, including that of Dijkema *et al.* and Konings *et al.*, could be consolidated using a so-called *bath-and-shower* effect. This effect, which is more well known to occur in spinal cord, causes the impact of radiation damage to a contained aspect of the parotid to be exacerbated by a lower dose to surrounding tissues [356]. van Luijk *et al.* demonstrated the effect in rat parotid and hypothesized that, analogously to spinal cord, enhanced radiation damage might be caused by depletion of stem/progenitor cell populations within excretory ducts. They also surmised that hampered cell replenishment would stymie recovery. Transplantation of stem cells have since been found to aid functional recovery [214, 360, 361].

Meanwhile, the primary aim of concurrent observational studies was to locate critical regions so as to alter the dose profile and naturally facilitate recovery and *prevent*, rather than *correct*, functional loss. Buettner *et al.* in 2012 reported that the use of dose distribution descriptors (moments) improved prediction of xerostomia over whole parotid mean dose [233]. Using moments, they demonstrated that a concentration of dose in the medial-caudal aspect is preferable to homogeneous delivery across the whole parotid with the same mean dose, demonstrating, *observationally*, a region-dependent bath-and-shower effect in human parotid. More recently, in 2015, Clark *et al.* employed sub-segmentation to partition ROI, including 'cleaving' segmentation with coronal and sagittal planes to generate volumetric halves [324]. As a whole, evidence of regional variations was negative, but, owing to the direct approach taken, function loss profiles were distinctly and significantly

related to the mean dose of individual parotid aspects. Shortly thereafter van Luijk *et al.* also applied planar segmentation, but instead used *axial* planes [357]. A confined critical region was identified near the middle-dorsal edge of the mandible. Focused dosimetric and histopathological work with animal parotids were used to support their finding, and an *in vitro* human sample was found to contain stem/progenitor cells in the vicinity of the critical area. Most recently, Miah *et al.* showed that bilateral sparing of the superficial lobe – not the edge adjacent to the dorsal edge of the mandible – will reduce severe xerostomia compared to the more conventional approach of sparing the whole contralateral parotid [358].

Recent findings have generated excitement (e.g., a recent article entitled "*Radiation-induced salivary hypofunction may become a thing of the past*" [437]). However, several issues remain. First, there is not yet consensus on the location of the most critical region(s). Second, varied or relatively small cohorts ($N < 100$) have been used to produce these findings. Third, disproportionate representation of parotid aspects or multiple comparison may have inadvertently been applied when comparing candidate regions. Finally, no interventional trials have thus far been reported; only data from animal and observational studies have been used. The latter shortcoming is especially significant because clinical best practices impose strong dose contraints on contralateral parotid. Coupled with the small spatial extent of parotid, dose profiles are therefore likely to exhibit strong inter- and intra-parotid correlation. Regressor correlation will confound relative importance analysis [438].

In this work, the theory that there are *dosimetrically critical sub-structures within the parotid that strongly influence production of saliva after radiotherapy* is tested. A single, large (N=332), prospectively-collected clinical cohort is analyzed. Instead of relying on direct comparison of correlate measures or model goodness-of-fit metric, relative importance of organ sub-segments are assessed using sensitivity analysis, explained variance methods, and importance drawn from ranking of candidate models that exclusively and proportionally represent individual sub-segments. Regressor correlation is anticipated and managed. Entire parotid volumes are segmented into sub-

segments with equal volumes instead of relying on any 'central' points or regions of varying extent. Sub-segment proportionality is ensured. Sub-segment extent does not depend on arbitrary parameters other than choice of segmentation, and segmentation is varied to minimize impact on our findings. No individual statistical metric is relied upon; multiple comparisons are avoided by applying importance methods in orthogonal domains, not repeating individual statistical tests, and with ranking, rather than absolute comparison. Finally, our analysis procedure was developed with a clinical focus so that the location of critical region(s) can be identified using only treatment planning ROIs, instead of requiring detailed knowledge of structure (e.g., lobes, fine anatomy, or functional information).

12.2 Materials and Methods

12.2.1 Cohort Selection, Quality Assurance, Dosimetric Extraction

This provisional study passed institutional ethical review. All patients gave informed consent. Stimulated saliva was measured prior to radiotherapy ('baseline'; W_b) and both 1y (W_{1y}) and 2y (W_{2y}) following treatment by measuring accumulated whole-mouth spittle stimulated by chewing unflavoured paraffin wax over 5 minutes. Patients were excluded if: $W_{1y}/W_b > 5$ (i.e., accounting for naturally low baseline and standard measurement variability); salivary glands were surgically removed; parotids had tumour involvement; or they received atypical chemotherapy agents, electron therapy, or previous interfering radiotherapy. Radiotherapy planning ROI contours were scrutinized by a single senior head and neck oncologist (JW). The entire cohort was collected within the BC Cancer agency.

Ipsilateral (i.e., 'hot') and contralateral (i.e., 'cold' or 'spared') parotid are differentiated by clinical guidelines. This distinction was carried forward into our analysis. Sub-segments with equal volume were generated using nested segmentation (n.b. discussed in chapter 11). Dosimetric data extraction and ROI manipulation were accomplished with `DICOMautomaton` [411].

12.2.2 Models

Models were fitted to saliva measurements via sub-segment mean dose regressors. Earlier work with non-parametric methods by Clark *et al.* demonstrated the dose-response shape for various sub-segments is primarily linear [324]. In this work, four prototypical linear models were considered. We demonstrate their form for 1/2-volume cranial-caudal segmentation. The first prototype ("lin-split") combines regressors without co-linkage except a shared intercept. The form is

$$\frac{W_{1y}}{W_b} = A - b_u^i M_u^i - b_u^c M_u^c - b_l^i M_l^i - b_l^c M_l^c. \tag{12.1}$$

A is the intercept, M_u^i refers to the mean dose delivered to the (u)pper (i.e., cranial) sub-segment of the (i)psilateral parotid, and b_l^c is the (c)ontralateral (l)ower (i.e., caudal) sub-segment slope. Ipsi- and contralateral sub-segments slopes are independent, permitting individual assessment. The second prototype ("lin-locked") unites regressors (left-right) providing combined ipsi- and contralateral parameters. The form is

$$\frac{W_{1y}}{W_b} = A + b_u \left(M_u^i + M_u^c\right) + b_l \left(M_l^i + M_l^c\right). \tag{12.2}$$

The third prototype provides sub-segment-exclusive variations of the "lin-split" type (e.g., cranial only: "lin-split-cranial"). The fourth provides sub-segment-exclusive variations of the "lin-locked" type. Sub-segment-exclusive model ranking provides relative importance derived from foundational information theory [439].

Exponential models are recommended for whole parotid in the literature (e.g., [26]). No specific recommendation for or against exponential models was found for sub-segments, so analogous exponential prototypes were included. The first prototype ("exp-split") is

$$\frac{W_{1y}}{W_b} = \frac{A}{4}\left(e^{-\alpha_u^i M_u^i} + e^{-\alpha_u^c M_u^c} + e^{-\alpha_l^i M_l^i} + e^{-\alpha_l^c M_l^c}\right). \tag{12.3}$$

Every regressor has a pseudo-slope. The factor A consumes the 'scaling'

degree of freedom consumed by the linear A. Exponential prototype #2 ("exp-locked") has the form

$$\frac{W_{1y}}{W_b} = \frac{A}{4}\left(e^{-\alpha_u M_u^i} + e^{-\alpha_u M_u^c} + e^{-\alpha_l M_l^i} + e^{-\alpha_l M_l^c}\right). \tag{12.4}$$

As with linear prototypes, the third and fourth exponential prototypes are sub-segment-exclusive. Exponential models are scaled such that A is unity under the null hypothesis.

12.2.3 Statistics

A standard statistical significance threshold (α) of 0.05 was used. Models are compared primarily on the basis of Akaike's Information Criterion (AIC) [440], but Mean Absolute Error (MAE) is also considered. AIC ranks models on the basis of an information-theoretic argument by accounting for both goodness-of-fit and model complexity. Akaike Weights (AW) are used to estimate the relative importance of groups of models (e.g., caudal vs. cranial, linear vs. exponential) [439]. Throughout, AW are used to report family-wise percentage-normalized Relative Importances (ARI) that describe the likelihood of the stated factor being most important. MAE and other related metrics, e.g., Root-Mean-Square Error (RMSE), eschew model complexity and characterize predictive power. There is debate over the appropriateness of MAE and RMSE for model comparison [441]. Both are common. MAE weights residuals equally, whereas RMSE gives higher weight to larger residuals, bolstering importance. Because whole-mouth salivary measurements are noisy, MAE was chosen. Model ranking performs dual function; identification of the best model(s) and factor importance assessment. Therefore, both AIC and MAE are considered when rejecting models.

To ensure an apples-to-apples comparison, linear and exponential models were fitted identically. A good model fit will have residuals that are not dependent on the regressors and distributed randomly (i.e., not systematically) about zero [442] when noise is normally distributed. However, baseline-normalized salivary measurements will not be normally distributed (nor homoscedastic) so residual normality was not tested. Instead, the non-

parametric 'runs' test was used to test for residual sign autocorrelation (i.e., detecting conspicuously consecutive *runs* of positive or negative residuals) [442].

Besides importance derived from AIC model ranks, fitted parameters are reported for representative models and compared (sensitivity analysis). A third class of methods that estimate 'dispersion' importance via ascribing fractions of the explained variance to individual regressors are also used [438]. Permutation methods (a.k.a. *averaging over orderings* methods) are represented by the 'LMG' measure of Lindeman, Merenda, and Gold [443]. Permutation methods are robust to multicollinearity, but are computationally demanding [438]. Another, simpler, less robust measure (referred to as 'LI') is employed as foil to LMG. It uses the increase in the coefficient of determination when including each regressor after all others are included. Since parotids have limited spatial extent, multicollinearity is anticipated. Spearman's rank correlation coefficients (ρ) are reported for adjacent and pairwise (left-right) sub-segment mean doses.

12.3 Results

12.3.1 Mean-scaling 2y Expectorate Measurements

This work focused on *late* dysfunction. It is not *a priori* clear what follow-up time is sufficient to capture late dysfunction, but based on section 6.7.2 salivary measurements appear to stabilize prior to the one year follow-up and remain approximately static, on average, until the two year follow-up. Many patients were unable to attend, or declined to attend either one year or two year follow-ups. We believe random factors contributed most significantly to patients missing follow-ups (e.g., patient commute being too far or challenging, recurrence leading to a second, interfering radiotherapy treatment), and so compared W_{1y} and W_{2y} distributions for equality. The hypothesis was that one year and two year follow-ups yield whole stimulated saliva measurements that can be considered drawn from the same base distribution. In other words, that patient-specific W_{1y} is identical to W_{2y} and can be used as surrogates for

one another. A Kolgomorov-Smirnov test showed that the distributions are not significantly distinct (two-sided $p = 0.12$), but the means ($\overline{W_{1y}} = 4.02$ g/5min and $\overline{W_{2y}} = 4.64$ g/5min) differed by 13% of W_{2y}. Scaling W_{2y} by $\overline{W_{1y}}/\overline{W_{2y}}$ also led to two significantly indistinct distributions (two-sided $p = 0.63$), but the means were identical. For reference, a comparison of W_b and W_{3m} yield a significant distinction (two-sided $p < 0.0001$), with means (7.00 g/5min and 3.02 g/5min, respectively) which differed by 132% of W_{3m}. Subtraction of W_{1y}/W_b and W_{2y}/W_b (for patients with both measurements only) resulted in a distribution that was not normal according to Shapiro-Wilk ($p < 0.001$), Anderson-Darling ($p < 0.001$), or Pearson's χ^2 ($p = 0.012$) normality tests. Subtraction of W_{1y}/W_b and mean-scaled W_{2y}/W_b achieved normality according to the Pearson's χ^2 normality test ($p = 0.077$) but not Shapiro-Wilk or Anderson-Darling tests. However, the distribution was centered near zero and evenly distributed with 57% of differences lying to the left of zero (see figure 12.1) so we believe the distribution is 'normal enough' for purposes of this work. The average W_{1y}/W_b was 0.584 if two year measurements were not used ($N = 303$), 0.580 ($N = 332$) if W_{2y} were substituted for missing W_{1y}, and 0.577 ($N = 332$) if mean-scaled W_{2y} were substituted. Based on these findings, mean-scaled W_{2y} were used as substitutes for missing W_{1y} as-needed in 29 cases. The total number of patients available for analysis was 332.

12.3.2 Distribution of Baseline-Normalized Salivary Measurements

Salivary measurements are not normally distributed. First, they cannot be negative. Second, they 'lean' toward zero. Normalization using the baseline measurement (e.g., W_{1y}/W_b) is a simple way to account for per-patient variability. It also, unfortunately, results in a heteroscedastic dependent variable.

The distribution of baseline-normalized salivary measurements is not *a priori* known. Using quantile-quantile plots [444] and the graphical method proposed by Cullen and Frey [9, 445], we have empirically determined that the distribution is approximately equal parts gamma and log-normal (i.e.,

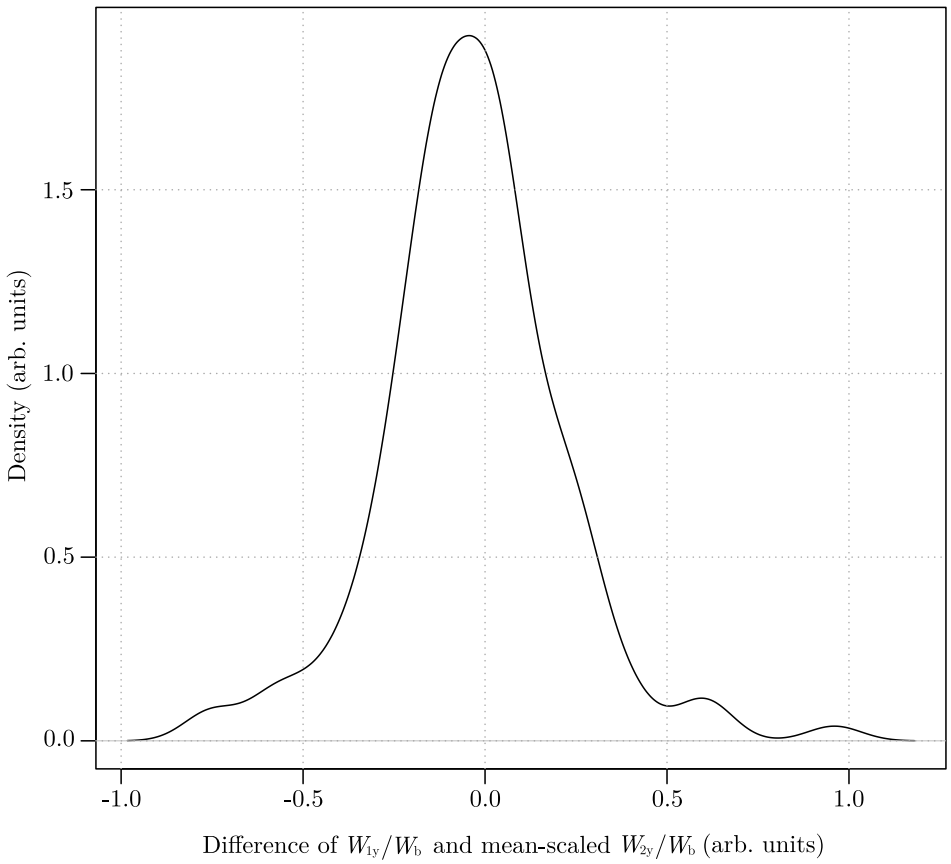


Figure 12.1: Kernel density estimate of the difference between W_{1y}/W_b and mean-scaled W_{2y}/W_b. The optimal bandwidth was estimated by the method of [8].

Weibull [446]; see fig. 12.2 and fig. 12.3.) Both distributions are amenable to maximum likelihood estimation via least-squares [447, 448]. While fitted parameter uncertainty estimates may be skewed or unreliable on account of non-normality of fitted model residuals, no importance is derived from these quantities and they are inconsequential to our analysis.

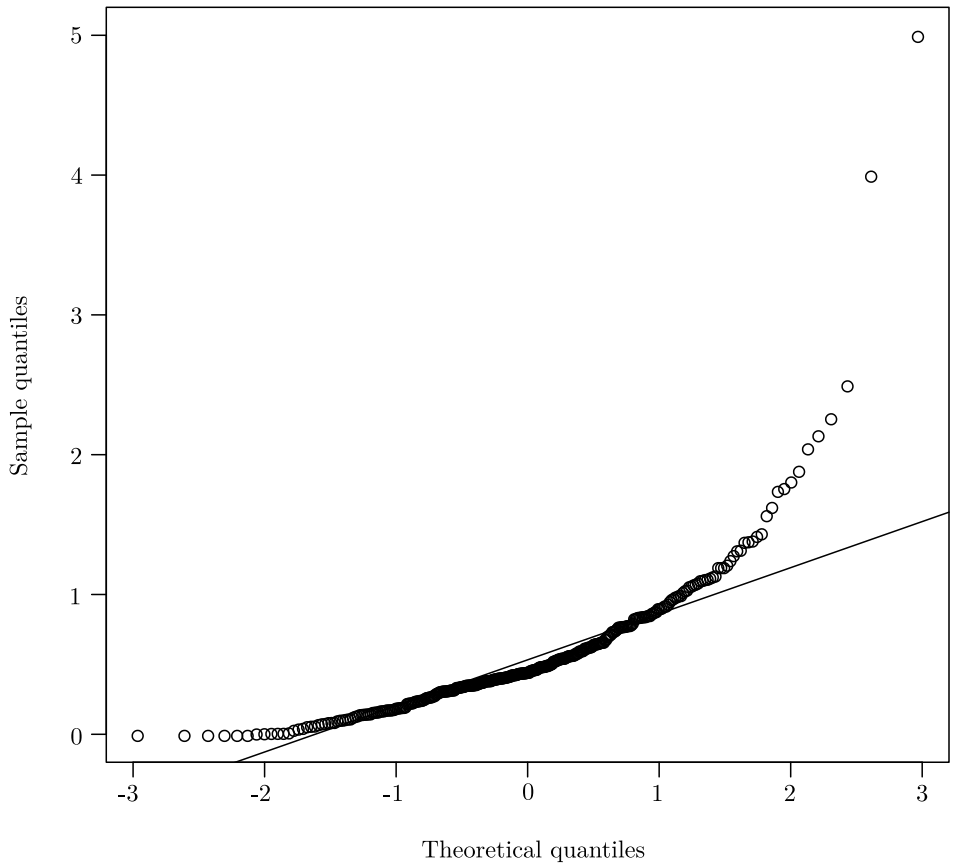


Figure 12.2: Quantile plot showing clear deviation from normality, but consistency with a gamma distribution.

12.3.3 Whole Parotid

Whole parotid sample mean doses for ipsi- and contralateral parotid were 30.50 ± 0.69 Gy and 16.67 ± 0.44 Gy respectively ($\pm$ std. dev.). Mean doses in this cohort were relatively ergodic. Figure 12.4 shows a scatterplot of the whole ipsi- and contralateral parotid mean doses. The region around the QUANTEC clinical limits can be seen to be more densely sampled than surrounding areas. Spearman's rank correlation coefficient between ipsi- and contralateral dose was 0.484.

Whole parotid mean dose was modeled to establish baseline AIC. The

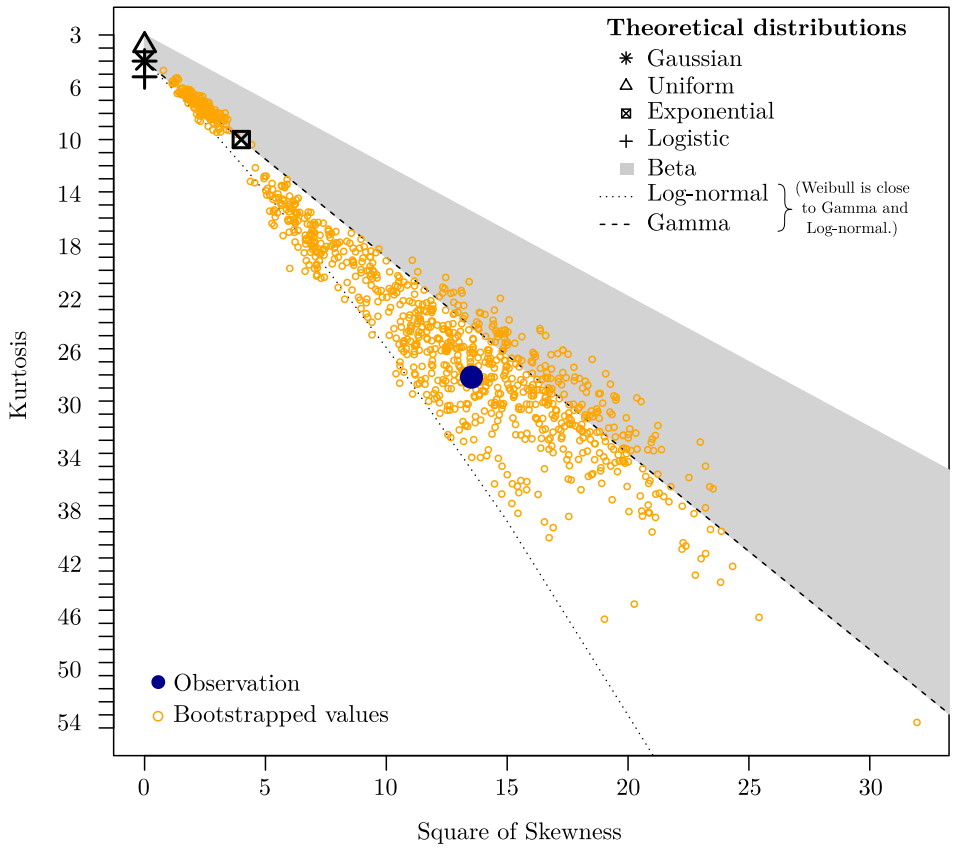


Figure 12.3: Distribution classification plot as proposed by Cullen and Frey [9]. 5000 bootstraps were performed. The empirical distribution is approximately equal parts gamma and log-normal.

linear model performed best, with AIC 469.60, compared to the exponential model (AIC +5.20; AW 0.07). Contralateral parotid dose-response slopes dominated ipsilateral in both linear (0.014 ± 0.003 Gy^{-1} vs. 0.0003 ± 0.0020 Gy^{-1}) and exponential models (0.044 ± 0.027 Gy^{-1} vs. 0.0061 ± 0.0066 Gy^{-1}). Intercepts spanned 0.84-0.86 and MAE was 0.30 in both cases. The proportion of variance explained by the linear model was 7.7%. Both LMG and LI measures showed the contralateral parotid to be substantially more important (both $>7.0\times$).

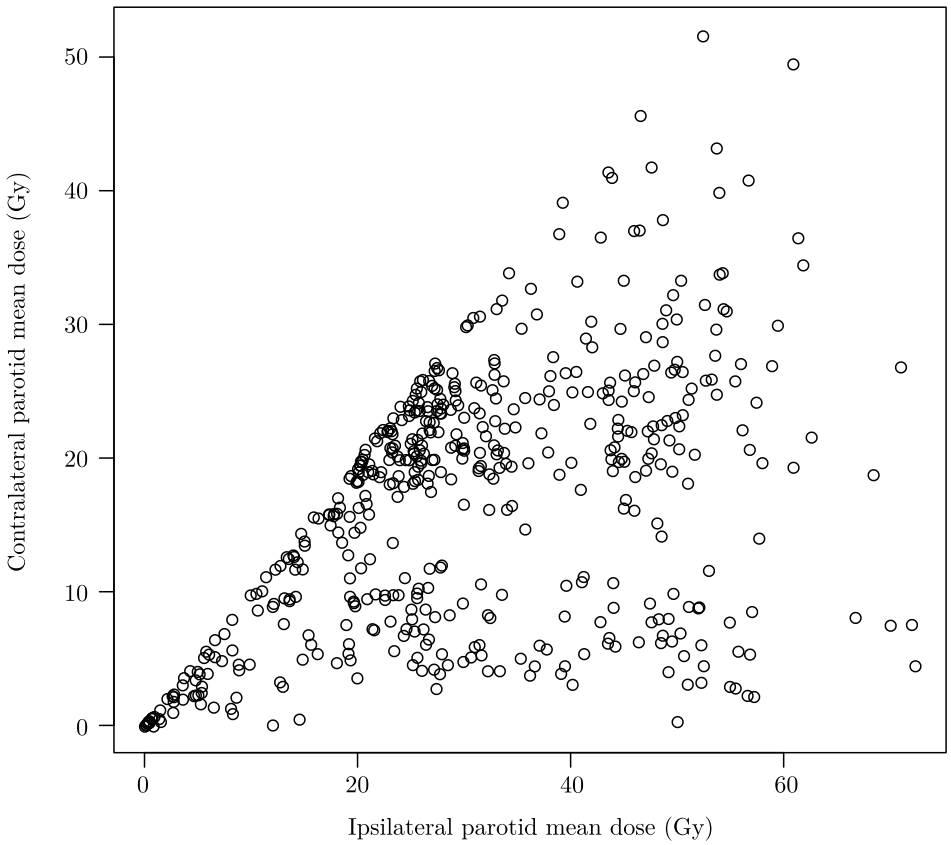


Figure 12.4: Scatterplot of whole ipsi- and contralateral parotid mean doses.

12.3.4 Cranial-Caudal 1/2-Volume Sub-Segments

Parotids were segmented into cranial and caudal halves. Sub-segment mean doses were significantly correlated when comparing adjacent sub-segments (i.e., cranial vs. caudal) in the ipsilateral parotid (Spearman's $\rho = 0.732$) and contralateral parotid (0.640) and also when comparing pairwise cranial (0.551) and caudal (0.469) sub-segments (two-sided $p < 0.0001$ in all cases). The average cranial sub-segment mean dose was 17.19 Gy (min 0.02 Gy, max 72.10 Gy) whereas it was 29.98 Gy (min 0.03 Gy, max 72.68 Gy) for caudal sub-segments; both had absolute percent differences of 25% with whole

parotid mean dose. In every patient at least one caudal sub-segment (i.e., left or right) received a higher mean dose than the adjacent cranial sub-segment.

Linear models fitted with the generic nonlinear regression framework were identical to those of ordinary least-squares. Table 12.1 shows model quality parameters. Fits converged without issue. Runs tests were not obviously correlated with either AIC nor MAE, but were highest for cranial-only and locked variant models. The overall best performing model in terms of AIC and MAE was lin-split (table 12.2). The '-split' variants outperformed '-locked' variants (AIC and MAE; ARI: 89.2% vs. 10.8%). Linear models outperformed exponential models (ARI: 70.7% vs. 29.3%). Caudal-only models performed better than their cranial-only counterparts in all cases and also family-wise (ARI: 82.1% vs. 17.9%); caudal-only model AIC was reduced by at least two compared with their cranial-only counterpart. Additionally, some of the lowest AIC models (n.b. those performing best) were caudal-only. Conversely, the three models with the highest AIC (n.b. those performing worst) exclusively featured the cranial sub-segment. MAE were similar for all models.

Model	DOF	MAE	AIC	AW
lin-split	327	0.30	472.99	0.308
lin-split-caudal	329	0.30	473.12	0.288
exp-split	327	0.30	475.15	0.104
exp-split-caudal	329	0.30	475.18	0.103
lin-split-cranial	329	0.30	475.79	0.076
exp-locked	329	0.30	477.28	0.036
exp-locked-caudal	330	0.30	477.51	0.032
lin-locked-caudal	330	0.31	478.56	0.019
lin-locked	329	0.31	479.27	0.013
exp-split-cranial	329	0.31	479.29	0.013
exp-locked-cranial	330	0.31	481.42	0.005
lin-locked-cranial	330	0.31	482.32	0.003

Table 12.1: AIC-ranked W_{1y}/W_b regression models using $^1/_2$-volume sub-segments. Models are ranked by AIC (lower is better). All quantities are dimensionless. AW denotes the Akaike weight. In all cases $p_{\text{runs}} > 0.14$.

Fitted parameters for the best whole-parotid models are shown in table 12.2. In all three models caudal sub-segments play a more dominant role than cranial sub-segments. Lin-split and exp-locked models provide the most compact slope estimates. In these models the caudal sub-segment slope was 1.5-3.6× larger than the cranial slope.

Model	Param.	Estimate	Std. Err.
lin-split	A	0.819	0.070
	b_u^i	0.0026	0.0034
	b_u^c	0.0056	0.0057
	b_l^i	−0.0020	0.0031
	b_l^c	0.0085	0.0037
exp-split	A	1.07	0.14
	α_u^i	0.004	0.010
	α_u^c	0.034	0.063
	α_l^i	0.41	0.88
	α_l^c	0.033	0.054
exp-locked	A	0.925	0.098
	α_u	0.0102	0.0093
	α_l	0.036	0.022

Table 12.2: Parameters for the best 1/2-volume W_{1y}/W_b whole-parotid regression models. All parameters except A have units Gy^{-1}; A is unitless. Superscripts denote (i)psi- and (c)ontralateral; subscripts denote (u)pper (cranial) and (l)ower (caudal) sub-segments.

The proportion of variance explained by the top model, lin-split, was 7.9%. If patients missing W_{1y} were excluded or mean scaling of W_{2y} was not used, the fitted model was slightly worse, explaining only 7.7% or 7.8% respectively. The relative importance of the four sub-segment mean dose regressors were estimated with explained variance measures. Contralateral parotid was more important than ipsilateral (combined: 5.8× by LMG, 6.3× by LI), and caudal sub-segments were more important than cranial (combined: 1.3× by LMG, 3.8× by LI). The single most important sub-segment was caudal (both LMG and LI).

12.3.5 Cranial-Caudal $^1/_3$-Volume Sub-Segments

Parotid were segmented into cranial, middle, and caudal thirds. Combined left-right sub-segment average mean dose for cranial was 15.48 Gy, for middle was 23.1 Gy, and for caudal was 32.7 Gy. In all cases extrema were <0.03 Gy and >72.11 Gy. Caudal and cranial sub-segment average mean doses had absolute percent differences spanning 31-49% with whole parotid. Middle sub-segments had percent differences of -0.2% and 0.1% (ipsi- and contralateral parotids, respectively), implying middle sub-segments are most representative of whole parotid mean dose. Adjacent sub-segment (i.e., cranial vs. middle) mean doses significantly correlated in ipsi- (Spearman's $\rho = 0.776$-0.828) and contralateral parotid (0.771-0.776); so did cranial (0.592), middle (0.472), and caudal (0.484) sub-segments (left-right parotid; $p < 0.0001$ in all cases).

Introduction of a third sub-segment slightly improved model quality (AIC) but not predictive power (MAE; see table 12.3). No runs tests were significant. Predictive power (MAE) across all models differed by <0.015. The best model was lin-split-caudal and again the best whole-parotid model was lin-split. Linear models were preferred over exponential models (ARI: 71.8% vs. 28.2%). Sub-segment models dispersed over the AIC spectrum; but family-wise caudal-only models had the best ARI (54.3%) followed by middle-only (40.0%) and cranial-only models (5.6%). Besides family-wise trends, individual sub-segment-exclusive models with the highest AIC (n.b. those performing worst) predominantly were cranial-only. The best performing models were caudal-only and to a lesser extent middle-only.

Fitted parameters for the best whole-parotid models are shown in table 12.4. Contralateral (pseudo-)slopes had the largest magnitudes in all but one case, and were more physically sensible than their ipsilateral counterparts. In lin-split and exp-locked models the caudal sub-segments play more dominant roles than cranial and middle sub-segments in the contralateral parotid ($b_l^c / \sup(b_u^c, b_m^c) > 2.0$ and $\alpha_l / \sup(\alpha_u, \alpha_m) > 2.6$). Again, the exp-split ipsilateral caudal pseudo-slope was dominant, but it was also anomalously large in magnitude. Contralateral pseudo-slopes, in contrast, were all of comparable magnitude.

Model	DOF	MAE	AIC	AW
lin-split-caudal	329	0.30	473.12	0.273
lin-split-middle	329	0.30	473.47	0.229
lin-split	325	0.30	474.24	0.156
exp-split-caudal	329	0.30	474.80	0.118
exp-split-middle	329	0.30	476.15	0.060
exp-locked	328	0.30	477.55	0.030
lin-split-cranial	329	0.31	477.63	0.029
exp-split	325	0.30	477.63	0.029
exp-locked-caudal	330	0.31	478.49	0.019
exp-locked-middle	330	0.30	479.00	0.014
lin-locked-caudal	330	0.31	479.52	0.011
lin-locked	328	0.31	479.53	0.011
exp-split-cranial	329	0.31	479.98	0.009
lin-locked-middle	330	0.31	480.62	0.006
exp-locked-cranial	330	0.31	481.57	0.004
lin-locked-cranial	330	0.31	482.70	0.002

Table 12.3: AIC-ranked W_{1y}/W_b regression models using $^1/_3$-volume sub-segments. All quantities are dimensionless. In all cases $p_{\text{runs}} > 0.18$.

Variance explained by the lin-split model was 8.62%. Despite an anomalous ipsilateral caudal slope, the *contra*-lateral caudal sub-segment was the most important individual regressor according to either metric. Importance ratios were >1.4 (LMG) and >6.8 (LI) compared to other sub-segments. Contralateral sub-segments were >3.1× more important than their ipsilateral counterparts in all cases (LMG). Combined LMG importances (percentage-normalized) were: 41.7% (caudal), 30.2% (middle), and 28.1% (cranial). LI importances were 83.3%, 0.7%, and 16.0%, respectively.

12.3.6 Cranial-Caudal $^1/_4$-Volume Sub-Segments

Parotid were segmented into cranial, middle-cranial, middle-caudal, and caudal quarters. Combined left-right sub-segment average mean dose was 14.33 Gy for cranial, 20.00 Gy for middle-cranial, 25.99 Gy for middle-caudal,

Model	Param.	Estimate	Std. Err.
lin-split	A	0.812	0.072
	b_u^i	0.0042	0.0047
	b_u^c	0.0043	0.0076
	b_m^i	−0.0011	0.0058
	b_m^c	0.0010	0.0076
	b_l^i	−0.0022	0.0035
	b_l^c	0.0085	0.0037
exp-locked	A	0.96	0.12
	α_u	0.032	0.043
	α_m	0.002	0.012
	α_l	0.080	0.090
exp-split	A	1.00	0.16
	α_u^i	0.023	0.080
	α_u^c	0.04	0.16
	α_m^i	−0.005	0.013
	α_m^c	0.04	0.18
	α_l^i	0.4	1.4
	α_l^c	0.04	0.16

Table 12.4: Parameters for the best 1/3-volume W_{1y}/W_b whole-parotid regression models. All parameters except A have units Gy^{-1}; A is unitless. Superscripts denote (i)psi- and (c)ontralateral; subscripts denote (u)pper (cranial), (m)iddle, and (l)ower (caudal) sub-segments.

and 34.11 Gy for caudal. In all cases extrema were <0.04 Gy and >72.16 Gy. Middle-caudal mean dose differed least from whole parotid (absolute percent difference 11% and 15%); middle-cranial spanned 11-16% and cranial/caudal spanned 35-57%. Adjacent sub-segment (i.e., cranial vs. middle-cranial) mean doses significantly correlated in ipsi- (Spearman's ρ = 0.863-0.881) and contralateral parotid (0.831-0.846); so did cranial (0.633), middle-cranial (0.499), middle-caudal (0.456), and caudal (0.490) sub-segments (left-right parotid; $p < 0.0001$ in all cases).

Introduction of a fourth sub-segment slightly improved model quality (AIC) but not predictive power (MAE; see table 12.6). The three best

performing models were all lin-split type. Linear models performed better overall compared with exponential models (ARI: 76.4% vs. 23.6%), though they were undifferentiated when ranked by MAE. Caudal sub-segments were more important than cranial sub-segments (ARI: 78.1% vs. 21.9%), but the middle-caudal sub-segment was itself most important (ARI: 46.7%), followed by caudal (31.4%), middle-cranial (19.7%), and cranial sub-segments (2.1%). Besides poor AIC, cranial and middle-cranial models also had the worst MAE. Residuals in all models had insignificant runs test (all $p > 0.16$).

Fitted parameters for the two best whole-parotid models are shown in table 12.6 (the third, exp-split, performed poorly). While the AIC of the third-best whole-parotid model (lin-locked) differed from the top model (lin-split) by 8.64, the difference was explainable mostly by lost degrees of freedom (+8). MAE of lin-locked models were worst. The exp-split model had a reasonable MAE but had a poor AIC. Ipsilateral parameters were greater in magnitude in all but one case. In the exp-locked model caudal sub-segments were dominant; in the lin-split model the middle-caudal ipsilateral sub-segment was the most dominant sub-segment, followed by the middle-cranial ipsilateral sub-segment.

The amount of variance explained by the lin-split model was 10.3%. Combined caudal sub-segments were most important using LI (38.4%), followed by middle-caudal (27.7%), middle-cranial (24.4%), and cranial (9.4%). Similarly, using LMG caudal sub-segments were most important (30.7%), followed by middle-caudal (27.5%), middle-cranial (23.6%), and cranial (18.1%; all percentage-normalized). The LMG metric also showed that each contralateral sub-segment was more important that their ipsilateral counterparts. The most important individual sub-segment was either caudal or middle-caudal (both LMG and LI).

12.4 Discussion

This work sought to locate sub-structures in the parotid gland that are dosimetrically critical for late loss of salivary function. Three different parametric methods were used to derive relative importance: explained

Model	DOF	MAE	AIC	AW
lin-split	323	0.30	471.94	0.414
lin-split-middle-caudal	329	0.30	474.02	0.146
lin-split-middle-cranial	329	0.30	475.00	0.090
exp-split-caudal	329	0.30	475.12	0.084
lin-split-caudal	329	0.30	475.21	0.081
exp-split-middle-caudal	329	0.30	475.36	0.075
exp-locked-middle-caudal	330	0.30	477.16	0.030
exp-split-middle-cranial	329	0.30	478.21	0.018
lin-locked-middle-caudal	330	0.31	478.67	0.014
exp-locked	327	0.30	479.24	0.011
exp-locked-caudal	330	0.31	479.67	0.009
lin-split-cranial	329	0.31	480.12	0.007
lin-locked	327	0.31	480.58	0.006
lin-locked-caudal	330	0.31	480.77	0.005
exp-split-cranial	329	0.31	481.79	0.003
exp-locked-middle-cranial	330	0.31	481.79	0.003
lin-locked-middle-cranial	330	0.31	483.00	0.002
exp-locked-cranial	330	0.32	483.25	0.001
exp-split	323	0.30	484.03	0.001
lin-locked-cranial	330	0.32	484.33	0.001

Table 12.5: AIC-ranked W_{1y}/W_b regression models using 1/4-volume sub-segments. All quantities are dimensionless. In all cases $p_{\text{runs}} > 0.16$.

variance, model ranking, and sensitivity analysis.

12.4.1 Model Fitting

Baseline-normalized whole-mouth stimulated saliva measurements were fitted with a variety of models via least-squares. The saliva measurement distribution can not be Gaussian because saliva measurements are non-negative, likewise baseline-normalized measurements are not Gaussian and are necessarily heteroscedastic. As our methods were parametric and based on maximum-likelihood estimates, verification of the assumptions required by least-squares was paramount. Inspection of the distribution of W_{1y}/W_b

Model	Param.	Estimate	Std. Err.
lin-split	A	0.792	0.072
	b_u^i	0.0087	0.0052
	b_u^c	−0.0013	0.0088
	b_{mu}^i	−0.0176	0.0074
	b_{mu}^c	0.0128	0.0098
	b_{ml}^i	0.0198	0.0073
	b_{ml}^c	−0.0085	0.0086
	b_l^i	−0.0097	0.0040
	b_l^c	0.0100	0.0042
exp-locked	A	0.99	0.13
	α_u	0.038	0.064
	α_{mu}	−0.004	0.011
	α_{ml}	0.031	0.041
	α_l	0.12	0.18

Table 12.6: Parameters for the best 1/4-volume W_{1y}/W_b whole-parotid regression models. All parameters have units Gy^{-1}; A is unitless. Superscripts denote (i)psi- and (c)ontralateral; subscripts denote (u)pper (cranial), (m)iddle-(u)pper, (m)iddle-(l)ower, and (l)ower (caudal) sub-segments.

showed an approximately log-normal or gamma distribution (section 12.3.2). For distributions in the exponential family (e.g., gamma), least-squares estimates are equivalent to maximum-likelihood estimates [447]. Likewise for heteroscedastic log-normal distributed data [448]. Despite the equivalency, both heteroscedasticity and multicollinearity can render least-square standard errors unsuitable for relative importance [438]. Robust regression methods can be used to correct for heteroscedasticity, but multicollinearity remains an issue and statistical efficiency is sacrificed. The three importance methods we have employed do not make use of least-square standard errors, and so use of least-squares is justified. Fitted model residuals are often tested for normality to assess goodness-of-fit. However, such tests are futile in this case since W_{1y}/W_b follow a strongly skewed distribution. A non-parametric test ('runs') was performed instead. No model rejected the null hypothesis

($p > 0.14$), and least-squares convergence was consistent and uneventful, so we believe maximum-likelihood estimates were achieved.

AIC cannot be compared when different data sets have been used. However, proportional segmentation is equivalent to model reconfiguration over the same data set (i.e., mean dose to whole parotid is equivalent to average mean dose of all sub-segments), thus enabling inter-segmentation comparison. Mean dose to whole parotid remained the best predictor of salivary flow in terms of AIC (469.60 vs. next-best 1/4-volume linear-split model with AIC +2.34), and was no worse in average prediction error (MAE 0.30). However, (1) the ΔAIC was too small to outright reject use of segmentation; (2) ignoring extra model parameter degrees of freedom shows the log-likelihood is substantially reduced via segmentation; and (3) refined segmentation resulted in greater explained variance (7.7% $\rightarrow$ 10.3%). Therefore, we cannot say definitely that use of segmentation improves salivary dysfunction prediction, but we can say it remains valid to use segmentation, and furthermore segmentation may indeed capture more detailed dose-response facets. Regardless of the predictive capacity, use of segmentation for deriving relative importance of spatial portions within the parotid appears justified.

12.4.2 Explained Variance Importance

Multicollinearity was prevalent; all adjacent and left-right pairs of sub-segments were correlated ($\rho > 0.456$, all $p < 0.0001$). The robust LMG method discovered by Lindeman, Merenda, and Gold in 1980 [443] is robust to multicollinearity and was therefore employed in our study. A major limitation of the method is that it is only applicable to the lin-split model. Conveniently, this was the best-performing whole-parotid model in every case. Therefore we believe LMG provides the strongest estimates of relative importance compared with the other two importance methods we have employed. Contralateral parotid was at least 3.1$\times$ more important than ipsilateral parotid. The caudal-most aspects of the parotid (i.e., the caudal-most 1/3-1/2 of the total volume) were found to be uniformly more important than cranial aspects for prediction of baseline-normalized salivary function. Caudal-most

aspects had at least 1.3× the importance of cranial-most aspects in every comparison. Caudal sub-segments in both ipsi- and contralateral parotid were more important than all other more cranial sub-segments in the same parotid, indicating a gradient of importance highest in the caudal aspect. In 1/4-volume segmentation, caudal and middle-caudal aspects were nearly of the same importance, with normalized importances differing by only 3.2%.

The LI method is not robust to multicollinearity [438]. However, the differences between LI and LMG measures were unremarkable. There was strong agreement that *caudal aspects and contralateral parotid were most important*.

12.4.3 Model Ranking Importance

Likelihood functions (opposed to maximum-likelihood model estimates) are not strongly affected by multicollinearity. Relative importance derived from AIC by family-wise factor comparisons originates from foundational information-theory and is thought to be reliable [439]. Sub-segment comparisons were performed on sub-segment-exclusive models. Sub-segment proportionality was ensured by balancing the number of regressors in each comparison. Differences in sub-segment-exclusive model AIC do not stem from differing model free parameter counts because all models have the same total number of free parameters. Therefore family-wise median AIC differences are the result of model performance and goodness-of-fit alone.

Significance testing of AW (e.g., via α-thresholding) is recommended against [439], so percentage-normalized probabilities of the relative importance of family-wise models are considered as-is. Joint comparison of all models showed linear models were favoured over exponential models (ARI: 80.2% vs. 19.8%). Sub-segment-exclusive models performed well compared to whole-parotid models, especially considering they had access to only 1/2, 1/3, or 1/4 the information available to whole-parotid models. In two cases, 1/3-volume lin-split-caudal and lin-split-middle, they outperformed whole-parotid models (AIC and MAE). Caudal aspects of the parotid were both individually and family-wise more important than cranial aspects (family-wise ARI: 82.9%

vs. 17.1%). The next most important sub-segment was almost always the immediately cranially-adjacent sub-segment, resulting in a caudal-cranial importance gradient. In a single case – 1/4-volume segmentation – the middle-caudal aspect was most important.

12.4.4 Sensitivity Analysis Importance

Sensitivity analysis is sufficient to determine relative dispersion importance in the absence of multicollinearity [438]. When present, results are skewed because correlated regressors become degenerate and it is difficult to separate the effects of each. Multicollinearity was prevalent in our analysis but the parotid's small spatial extent caused it to be fairly consistent ($\rho \in [0.456, 0.633]$ inter-parotid, $[0.640, 0.881]$ intra-parotid). We therefore believe sensitivity analysis to be useful as a complementary technique.

Caudal sub-segments were most important. For 1/2- and 1/3-volume segmentation, caudal sub-segments had 1.5× or greater the importance of cranial sub-segments. A caudal-cranial importance gradient was again noted. In 1/4-volume segmentation, caudal-most sub-segments were more important than middle sub-segments in all but one model. In that exception, the middle-caudal sub-segment was most important.

12.4.5 Overall Assessment and Comparison with Earlier Studies

Combining results from the three methods, it is clear that caudal aspects of the parotid are most important for describing salivary performance one year after radiotherapy. In every case an importance gradient was noted with caudal aspects most important and cranial least important. In 1/4-volume segmentation, all three methods found that the middle-caudal sub-segment had comparable or greater importance compared with the caudal sub-segment. This finding is spatially consistent with 1/2- and 1/3-volume segmentation if the region encompassing 15-20% of the parotid volume, offset from the caudal-most aspect by ∼20% of the volume, contains the most important aspects. We find no reason to exclude this possibility.

Consensus about the existence and location of dosimetrically critical regions within the parotid gland has not yet been established in the literature. Konings *et al.* reported that salivary dysfunction in rat parotid was more pronounced after irradiation of the cranial aspect and that the cranial aspect seemed to impact the (unirradiated) caudal aspect [348]. The bath-and-shower effect reported by van Luijk *et al.* made use of a 30 Gy 'shower' dose to the caudal aspect with varying 0-10 Gy 'bath' doses to the cranial [356]. They found that addition of a dose bath resulted in increased dysfunction. However, guided by the stem/progenitor cell hypothesis, follow-up work by the same group identified a confined critical region within the *cranial* aspect [357] (n.b. 74 patients are shared between our analyses). Collectively, our findings do not appear consistent. While the discrepancy with rat parotid findings might arise from greater control over the rat parotid dose profile or differences in rat and human anatomy. Additional analytical factors may also play a role. For example the selection of critical region(s) by van Luijk *et al.* was accomplished with a spatially-variable bounding method. Our analysis used sub-segments with fixed volume and position, and importance of all sub-segments was quantified simultaneously. Our analysis differentiated ipsi- and contra-lateral parotids. Omitting this potentially confounding factor would have confined our analysis to the '-locked' model variants, which underperformed compared to '-split' variants. It could also reduce precision and give rise to a Yule-Simpson effect [385]. Despite differences, our conclusions on human parotid may broadly agree with those of van Luijk *et al.* since the middle-dorsal and

Buettner *et al.* found that dose distribution descriptors improved prediction over whole parotid mean dose, and noticed that dose concentrated in the medial-caudal aspect was preferable to homogeneous delivery across the whole parotid with equivalent mean dose [233]. In light of our findings, it may be that the observed effect is not an example of a homogeneous bath-and-shower effect but rather that they have singled out the most important aspect specifically (medial-caudal). Miah *et al.* showed that bilateral sparing of the superficial lobe reduces severe xerostomia incidence compared with conventional whole contralateral parotid sparing [358]. Laterality was not considered in our analysis, but our collective findings appear congruent if

the caudal aspects of superficial lobe are indeed the origin of the clinical response. Congruence in this case seems likely since, based on our findings, the caudal aspects of the superficial lobe may be most important. Finally, it is interesting that our earlier work was not able to detect significantly different dose-response in medial-lateral aspects. Given that the importance gradients observed in the present work are caudal-cranial, it seems likely that caudal-cranial tissue differentiation is more important than medial-lateral or anterior-posterior differentiation.

Segmentation refinements, such as incorporating additional coronal or sagittal planes, may improve localization of the critical region(s). We did not consider refinements for four reasons. First, radiotherapy is known to shrink parotid volumes by medial movement of the lateral aspects [305]. The present method is robust to this shrinkage. Second, a parametric approach quickly becomes untenable as regressors are added, which limits segmentation refinement. Third, refinement amplifies relative noise for individual sub-segments. Finally, refinement increases the difficulty of ensuring sub-segment volume proportionality.

12.4.6 Implications and Limitations

Zero parotid dose should result in no changes in function, so A=1 should hold in all considered models. Though linear models performed best, exponential model pseudo-intercepts (A) were closer to unity. They therefore may represent the data in a more physically sensible way. Pinning exponential A=1 did not alter model ranks; linear still performed best. Furthermore linear models always comprised the top 2-3 models – even whole parotid favoured linear models. It is possible that measurement noise, even with N=332, obscures fine details needed to distinguish models. Therefore linear models are recommended for cohorts approximately $\leq$500. This is contrary to the general consensus for whole parotid (i.e., typically exponential) but not uncommon. To our knowledge this work is the first to compare models for parotid sub-volumes in this way.

Middle sub-segment mean doses were most representative of whole parotid

mean doses. If parotids are truly homogeneous organs, and whole parotid mean dose is the best predictor, then middle sub-segments may have inflated importance. Middle sub-segments were important, but caudal sub-segments were generally more so. Since caudal sub-segments were least representative of whole parotid, their importance over middle sub-segments is noteworthy. Conversely, given that the amount of variance explained was 10.3% or less, and studies have shown the intensity of salivary gland damage and dysfunction increases in proportion to the irradiated volume [232], it seems likely that no single, small critical region exists that substantially controls whole-gland function. Broad caudal sub-segments might contain bulky critical regions. Refined segmentation methods are needed to test this hypothesis.

Finally, we emphasize this was an observational study. Treatments followed clinical guidelines and salivary dysfunction was aggressively minimized, which resulted in a relatively homogeneous (i.e., non-ergodic) cohort. Sub-segment importances may merely reflect the clinical dose profile, though we have endeavoured to overcome this tautological conclusion. It is therefore difficult to generalize our findings or ascribe radiobiologic significance. Interventional studies are needed to establish generalizability. On the other hand, our findings should be applicable when current clinical guidelines are followed [236], in which case the caudal aspects of the parotid should be spared as much as possible to ameliorate radiation induced dysfunction.

Chapter 13

Non-Parametric Approach to Regional Effect Assessment[1]

13.1 Introduction

Whole parotid mean radiation dose is currently used to predict risk of late radiotherapy-induced salivary dysfunction [236]. The underlying assumption is that functional burden is distributed homogeneously throughout the parotid gland [285]. Recent studies have found behaviour counter to homogeneous distribution, including regions with elevated relevance for salivary flow [324, 357], non-equivalence of dose-volume descriptors for dysfunction prediction [449, 450], and bath-and-shower effects [233, 356]. Others have noted that incorporation of non-homogeneous effects into a radiotherapy treatment plan leads or potentially could lead to improved patient outcomes [358, 428].

Evidence of a bath-and-shower effect in parotid, in which high dose to a confined sub-volume (the 'shower') is impacted by a low dose to an extended volume (the 'bath'), was first reported by van Luijk *et al.* in 2009 in the context of objective salivary flow dysfunction [356]. A similar effect was found

[1]The contents of this chapter have been submitted to *Radiation Oncology* under the title '*Fine segmentation shows anterior-caudal parotid is most important for salivary loss*' on the 31st of December, 2016. The list of authors was Haley Clark, Steven Thomas, Stefan Reinsberg, Allan Hovan, Vitali Moiseenko, and Jonn Wu.

using a separate cohort and subjective measurements in 2012 [233]. Likewise, several studies have confirmed that dose-volume measures are not equivalent in parotid, implying deviation from homogeneity. For example, Ortholan *et al.* found that salivary flow prediction improved compared to whole mean dose models when the volume of the contralateral gland receiving $\geq$ 40Gy was incorporated [449]. Wang *et al.* found similar conclusions in 2011 [450]. However, neither dose-volume effect deviations nor bath-and-shower effects incorporate *specific* sub-volumes; incorporation of sub-volume extent and location has lead to less conclusive findings. There is continued debate over the existence of critical regions (i.e., defined by specific anatomical, functional, or geographical criteria) that more strongly impact salivary dysfunction than comparable regions in the parotid. Different studies have variously shown that the most important regions are (or contain, or are contained broadly within) cranial and medial-dorsal aspects adjacent to mandible [348, 357], caudal aspects[2], caudal-medial aspects [233], the superficial lobe (i.e., approximately lateral-caudal) [358], and the lateral portion [324]. Other work has focused on the clinical feasibility of split delineation along the deep-superficial lobe boundary (i.e., anterolateral and posteromedial) [428, 451].

In this prospective study a cohort comprised of 332 head-and-neck cancer patients (collected within a single agency) is used to assess regional effects within parotid gland. Parotids are divided into 2, 3, 4, 18, and 96 equal-volume sub-segments. Sub-segment relative importance for prediction of late salivary flow is assessed using non-parametric methods robust to overfitting and multicollinearity. Owing to the linear dose-response observed in sub-segments in chapter 12 and in some capacity within the literature (e.g., [374]), and sub-segment volumetric equality, importances are interpretable as regional criticality for late salivary dysfunction.

[2] See chapter 12.

13.2 Materials and Methods

13.2.1 Cohort, Measurements, Treatment, Tooling

This prospective study passed institutional ethical review. Patients underwent radiotherapy for head-and-neck cancers and gave informed consent to participate. Planning dose profiles and delineated organ-at-risk parotid contours were employed for dosimetric assessment and segmentation. A single senior head and neck oncologist (JW) scrutinized contours for quality assurance. Stimulated late salivary measurements of whole-mouth saliva at baseline (pre-radiotherapy; W_b) and one year post-radiotherapy ("late"; W_{1y}) were used. Measurements represent whole mouth expectorate collected over five minutes of chewing flavourless wax. Mean-scaling imputation was employed for (29) patients without W_{1y} but with W_{2y} late measurements. Exclusion criteria are the same as in section 12.2.1. A total of 332 patients were eligible (median age 58.6y, age range 19.0-90.6y; gender: 73% male, 27% female; prescription dose: 70Gy/35 fractions 55%, 60Gy/25 fractions 11%, 60Gy/35 fractions 8%, other 27%; treatment type: 279 intensity- or volumetric-modulated, 53 conventional; primary tumour site: 88 nasopharynx, 132 oropharynx, 61 tongue, 61 tonsil, 31 oral cavity and gums, 20 unknown, 18 hypopharynx, 14 larynx, 7 thyroid, 4 palate, and 22 other).

Dosimetric accumulation and contour manipulation (nested segmentation) was accomplished via `DICOMautomaton` [7, 411] in accordance with the method of chapter 11. To ensure sub-segment proportionality, cubic dose matrix supersampling (15×) was employed. Counts of supersampled voxels within sub-segments were compared to ensure mutual pairwise proportionality using Kolmogorov-Smirnov tests. Significance was ascribed at α=0.05. No correction was made for multiple comparisons (i.e., to account for the so-called birthday paradox), which made for a *more stringent* test.

13.2.2 Importance Techniques

The Random Forest technique (RF) is a non-parametric ensemble learning method in which tree nodes are recursively constructed by randomly sampling

regressors at, and splitting, each node. An ensemble of trees is grown; regression predictions are generated by averaging predictions from the ensemble. Importance was estimated using two measures: (1) ensemble-averaged total decrease in node impurities resulting from splitting on the regressor and measuring the Residual Sum of Squares (RSS) ("node impurity"), and (2) a more robust permutation-based measure in which the difference between un-permuted and each regressor permutation of the out-of-bag (i.e., excluded data) Mean Squared Error (MSE) is ensemble-averaged and normalized by the standard deviation of the differences (referred to as simply "MSE" here) [452, 453, 454]. Major weaknesses of RF arise when regressors have varying scales, are mutually correlated ('multicollinearity'), or when the 'scale' (i.e., number of categories) of categorical variables differ [455]. In the present case all regressors (i.e., sub-segment mean doses) have the same scale and are continuous. Multicollinearity is anticipated, but is believed to be sufficiently pervasive and constant so as to reduce impact on conclusions by uniformly suppressing absolute importances and leaving relative importances intact.

RF trees may nonetheless become biased. To overcome this, conditional inference tree ensembles ("c-trees") were employed [456]. Like RF, c-trees can be used for non-parametric regression [457]. C-tree methods differ from RF by using conditional inference trees as base learners. The unbiased c-tree RF construction proposed by [455] is used, which is meant to address regressor selection bias in individual classification trees. Regressor importance is estimated using both (1) permutation and (2) conditional permutation measures. The former is a reliable measure of regressor importance for uncorrelated regressors when subsampling without replacement and unbiased trees are used to build the forest [455]. The latter, conditional permutation, is thought to be more suitable in the presence of multicollinearity and addresses regressor selection bias in individual classification trees [458].

Both RF and c-trees are thought to be robust to overfitting due to use of *bagging*, the assemblage of many bootstrapped trees and use of prediction averaging, which improves generalizability [452, 457]. Based on expected multicollinearity, the reliability of importance estimates were ranked as: c-tree conditional permutation (most reliable), c-tree permutation, RF MSE, and RF

node impurity (least reliable). The number of trees and splitting parameter were grown until impact on importances diminished and the random seed had no impact on conclusions (nominally 20k for both RF and c-trees).

13.2.3 Statistics

AIC is typically used to compare (parametric) models [459]. Besides an asymptotic relationship between cross-validation and AIC [460], the authors are not aware of any direct way to compute AIC for RF or c-trees. Instead, a metric that characterizes predictive power is used. Both MAE and RMSE [441, 461]. Both are reported. Fitted whole parotid mean dose models (linear and exponential; standard in the literature) provide baseline MAE and RMSE. The distribution of baseline-normalized salivary measurements will be heteroscedastic, so residual normality was not tested. Instead correlation coefficients (r_{pa}) between predicted and actual W_{1y}/W_b, are reported. Comparison is accomplished via a two-tailed Fischer z-transformation [462].

13.3 Results

A summary of all models and methods is shown in table 13.1. Contralateral parotid was unanimously more important than ipsilateral parotid for segmentation into halves, thirds, and quarters. Therefore, to reduce computational burden, segmentation into 18^{ths} and 96^{ths} used only contralateral parotids. MAE, RMSE, r_{pa}, and summarized importances are shown where applicable. C-tree methods performed significantly better than whole parotid mean dose models and RF (linear and exponential; both $p < 0.0001$) at all segmentation levels. RF methods did not significantly improve prediction when segmentation was introduced ($p \geq 0.258$) but c-trees improvement significantly strengthened ($p < 0.039$), improving from a correlation that was already nearly *double* the next-best method (0.531; linear model). Refinement-induced reductions in both MAE and RMSE were similar for RF and c-tree methods (ΔMAE: -0.013 vs. -0.016; ΔRMSE: -0.22 vs. -0.20). At all levels of segmentation a Kolmogorov-Smirnov test showed no statistically significant differences between the number of supersampled dose matrix voxels contained within each sub-segment ($p > 0.05$ in all $(96{\cdot}95+18{\cdot}17+4{\cdot}3+3{\cdot}2+2{\cdot}1)/2 = 4723$ comparisons).

Segmentation	Method	MAE	RMSE	r_{pa}	Type	Most Important Sub-segment	Importance
Whole	exp	0.301	0.491	0.252	–	–	–
	linear	0.295	0.487	0.277	–	–	–
	RF	0.315	0.506	0.222	–	–	–
	c-trees	0.259	0.437	0.531	–	–	–
Halves	RF	0.294	0.488	0.272	Impurity	caudal (contralateral)	1.15×
					MSE	caudal (contralateral)	1.45×
	c-trees	0.246	0.425	0.591	Permutation	caudal (contralateral)	2.78×
					Conditional	caudal (contralateral)	2.66×
Thirds	RF	0.308	0.494	0.246	Impurity	caudal (contralateral)	1.31×
					MSE	caudal (contralateral)	1.72×
	c-trees	0.249	0.422	0.611	Permutation	caudal (contralateral)	3.49×
					Conditional	caudal (contralateral)	3.05×
Quarters	RF	0.306	0.498	0.228	Impurity	middle-caudal (contralateral)	1.29×
					MSE	caudal (contralateral)	1.55×
	c-trees	0.247	0.421	0.614	Permutation	caudal (contralateral)	3.25×
					Conditional	caudal (contralateral)	2.70×
18^{ths}	RF	0.306	0.489	0.276	Impurity	SS04: caudal-anterior	1.47×
					MSE	SS14: middle-posterior	1.42×
	c-trees	0.248	0.420	0.620	Permutation	SS04: caudal-anterior	2.74×
					Conditional	SS04: caudal-anterior	3.85×
96^{ths}	RF	0.302	0.484	0.304	Impurity	SS04: caudal-anterior	2.47×
					MSE	SS26: middle-caudal-anterior	1.78×
	c-trees	0.243	0.417	0.637	Permutation	SS21: caudal-posterior	3.75×
					Conditional	SS04: caudal-anterior	4.04×

Table 13.1: Summary of results and most importance sub-segments. All quantities are dimensionless. r_{pa} denotes the correlation coefficient between actual and predicted mean-scaled W_{1y}/W_b. Whole, halves, thirds, and quarters segmentation used both ipsi- and contralateral parotids; 18^{ths} and 96^{ths} used only contralateral parotids to reduce computational burden. The most important sub-segment (SS) is specified; refer to fig. 13.3 for sub-segment locations. Importances given are relative to the expected result for a homogeneous parotid.

In almost every importance assessment method, a caudal-most sub-segment was most important. In the two exceptions, the most important sub-segment (middle; between caudal-most and cranial-most sub-segments) was either fully or partially within the caudal 50%-volume. In one of these exceptions, the 18^{ths} segmentation RF-MSE case, the next most important non-middle sub-segment was caudal.

The most important sub-segments, on average over all segmentation methods, had importances 2.4× that of an equivalent sub-segment in a theoretical homogeneous parotid (see table 13.1). This figure increased when segmentation and methodology was refined: 2.7× when only 18^{ths} and 96^{ths} segmentation was considered, 3.0× when only 96^{ths} segmentation was considered, and 4.0× when only c-tree conditional permutation (the most reliable method) was considered at the finest (96^{ths}) segmentation.

Other than the most important sub-segment, the least important sub-segment, median importances of family-wise groupings based on anatomy (e.g., caudal vs. middle vs. cranial, or anterior vs. posterior), and family-wise percentiles (e.g., 20% and 80%) conveyed similarly the importance of caudal aspects. Tables showing sub-segment importance of the most reliable technique, c-tree conditional permutation importance, are shown in figs. 13.1 and 13.2. The same information is displayed in the form of heat maps for 18^{ths} and 96^{ths} segmentation in fig. 13.3.

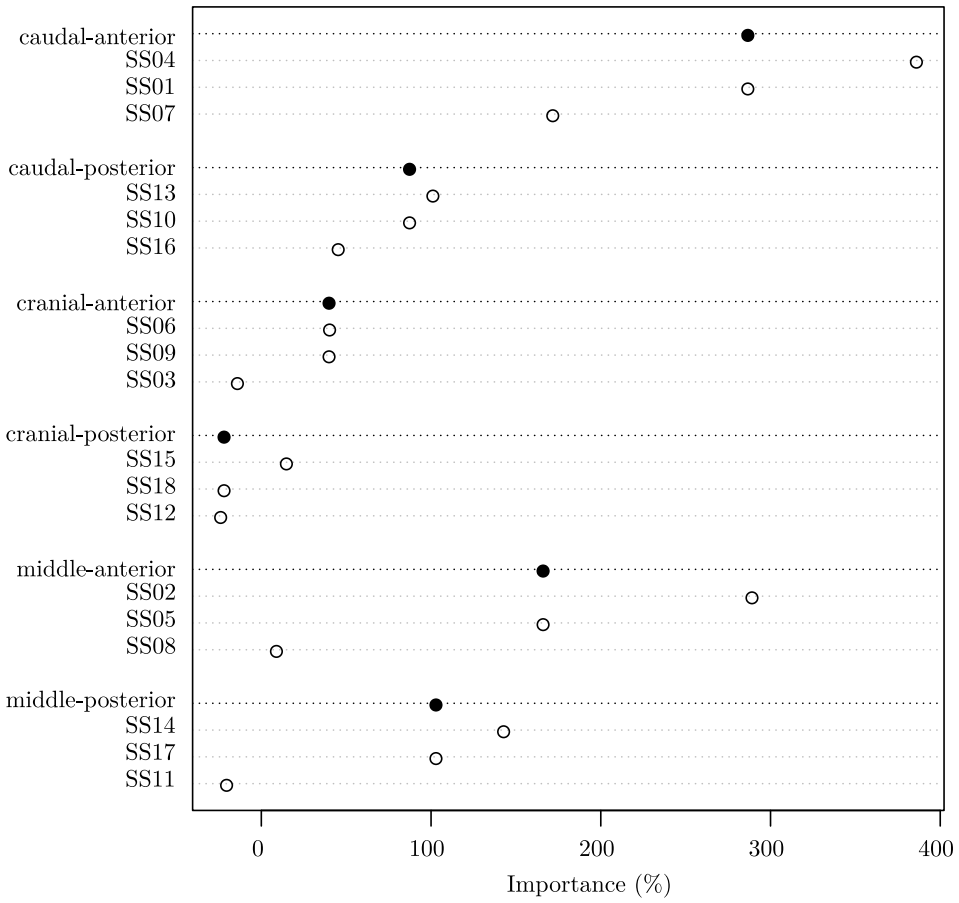


Figure 13.1: Relative c-tree conditional permutation importance of sub-segments for 18^{ths} segmentation. Importance is given as the percentage of relative importance compared to a homogeneous organ (which would be 100%). Refer to fig. 13.3 for sub-segment ('SS') spatial correspondence. Anatomical groupings display the per-group median (filled circles). Importances span ∼0-3.85× that of equivalent sub-segments in a homogeneous parotid.

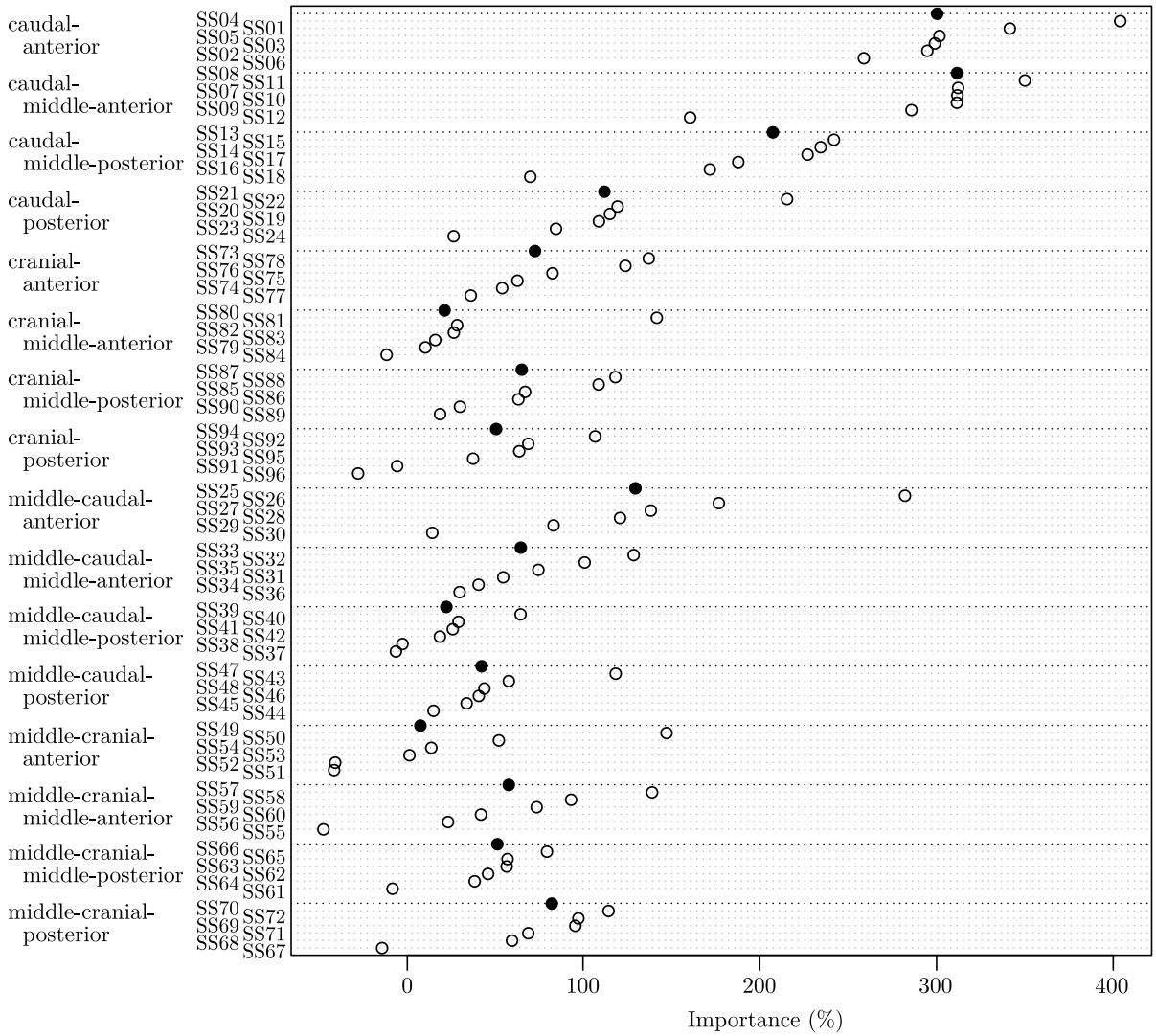


Figure 13.2: Relative c-tree conditional permutation importance of sub-segments for 96^{ths} segmentation. Importance is given as the percentage of relative importance compared to a homogeneous organ (which would be 100%). Refer to fig. 13.3 for sub-segment ('SS') spatial correspondence. Anatomical groupings display the per-group median (filled circles). Importances span ∼0-4.04× that of equivalent sub-segments in a homogeneous parotid.

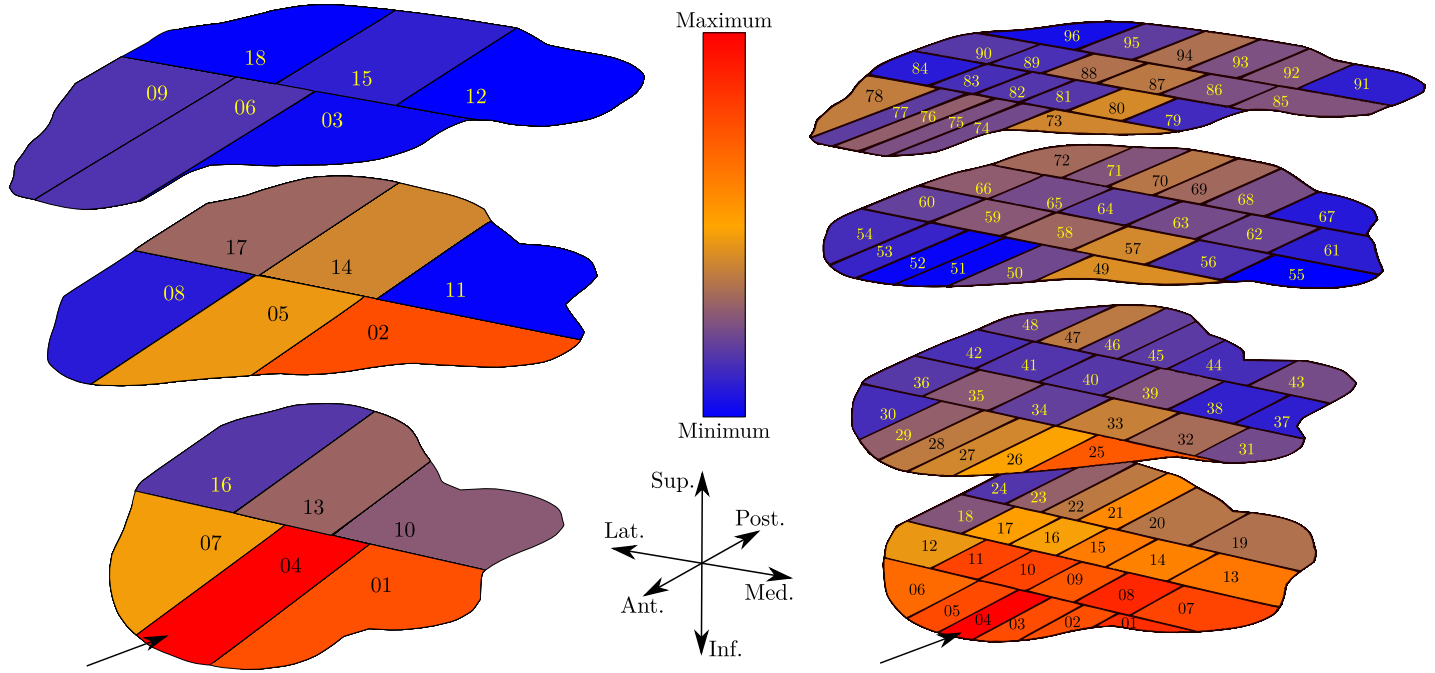


Figure 13.3: Relative c-tree conditional permutation importance of sub-segments for 18^{ths} (left) and 96^{ths} (right) segmentation. Equal-volume sub-segments are represented by a single slice of axial plane encompassed by the sub-segment. In segmentation into 18^{ths} (96^{ths}), importances span ∼0-3.85× (∼0-4.04×, respectively) that of equivalent sub-segments in a homogeneous parotid. The most important sub-segments are indicated.

Figure 13.4 shows contralateral parotid dosimetric characteristics for the cohort. Dose was highest in the caudal-medial aspects and lowest in the cranial-lateral aspects, on average. Importances did not merely follow sub-segment mean dose or dosimetric variability throughout the sample population (cf. figs. 13.3 and 13.4).

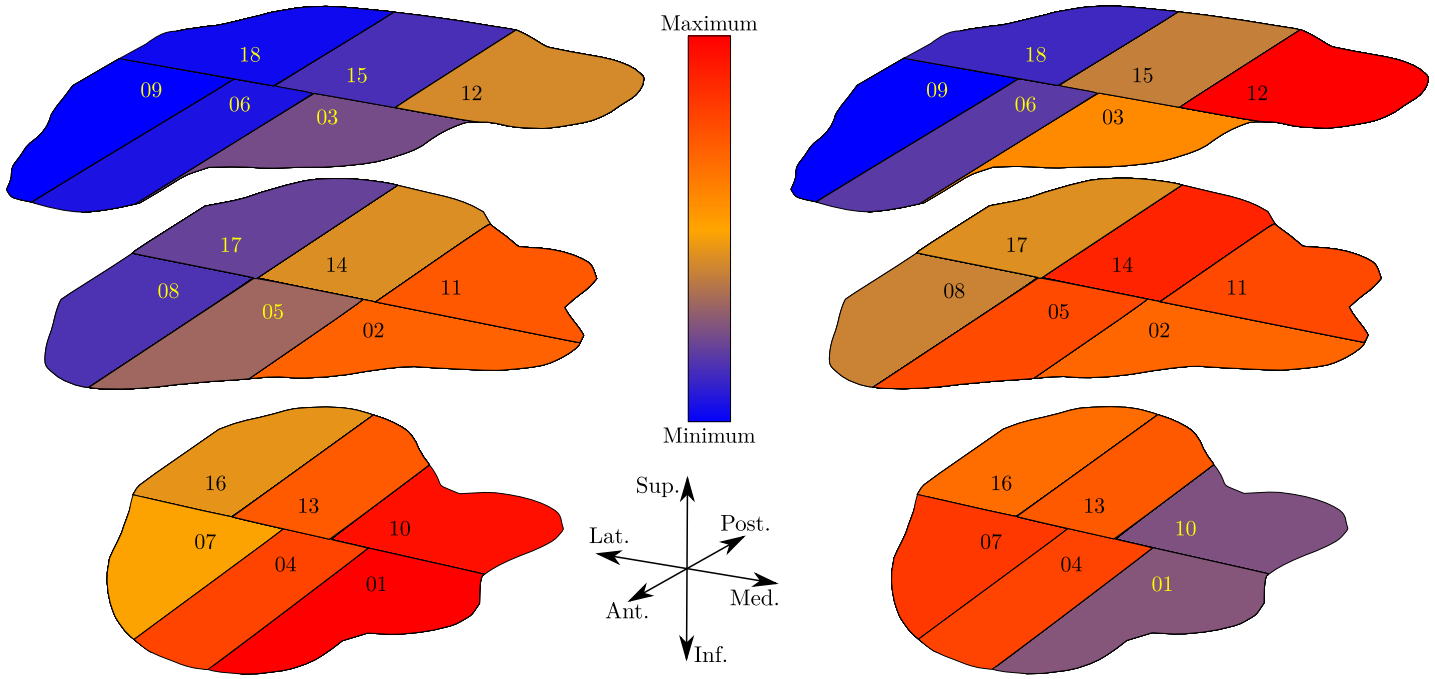


Figure 13.4: Sample population contralateral parotid dosimetric characteristics: mean dose (left) and the inner-most 50th percentile of mean dose (right) for each sub-segment (SS). Mean doses span 15.4-50.2Gy (SS09 and SS01, respectively). Inner 50th percentiles span 16.1-33.9Gy. Caudal-medial aspects received the highest dose while cranial-lateral aspects received the lowest dose, both with low variation across the sample population.

13.4 Discussion

Effects that deviate from strict parotid functional-spatial homogeneity have been reported, but there is not yet consensus about the criticality of specific sub-volumes in relation to radiotherapy-induced salivary dysfunction. In this work, a regional effect is characterized via a segmentation refinement method. We improve upon existing studies primarily by being systematic in coverage of the parotid: no aspects were *a priori* omitted and importance of the whole parotid is simultaneously developed.

Four non-parametric methods were used in this work. Though they varied in susceptibility to multicollinearity and other biases, all confirmed the importance of caudal aspects for predicting radiotherapy-induced late salivary function. Contralateral parotids were found to be most important, which is consistent with much of the literature (e.g., [449]). Sub-segment heat maps overlapped across segmentations and importance methods, which suggests conclusions do not substantially depend on the spatial resolution or other convergence factors (e.g., number of trees). A gradient of importance emerged indicating both caudal-anterior aspects are most important and that importance gradually fades posteriorly and superiorly. Starting at the most important sub-segment, movement to superiorly-adjacent regions affected the greatest reduction in importance. Posterior movement less so, and medial and lateral movement affected importance only weakly and approximately equally. Lack of medial-lateral preference may result from parotid medial shrinkage during radiotherapy [74]; lateral aspects may have traveled medially and 'smeared' importance. It remains to be seen if this effect is a treatment artifact.

C-tree methods outperformed RF significantly, and while they both generally improved MAE and RMSE as segmentation proceeded, only the c-tree r_{pa} significantly improved. It is not possible to ascribe this to any specific factor, but it is likely that either (1) RF is intrinsically not capable of ferreting out the information that an equivalent c-tree ensemble can, or (2) RF was strongly impacted by multicollinearity or measurement noise and tree construction was biased. In either case, while RF did not significantly

perform *better* than whole parotid models, it was also not significantly *worse*, and we therefore believe it remains a valid tool for inspecting sub-segment importance.

Though there is general consensus among researchers that the parotid is not homogeneous, there is little consensus about the specifics of the inhomogeneities. The existence of critical regions, mechanisms supporting them, and comparative clinical relevance of various aspects and lobes are debated. The region we have found to be most important overlaps, at least somewhat, with critical regions reported in previous studies. Buettner *et al.* in 2012 compared the relative importance of 50 clinical and physical factors (both categorical and continuous) for subjective xerostomia in 63 head-and-neck cancer patients [233]. Four of the seven most important regressors (mean dose to either parotid, contralateral parotid caudal-medial aspect dose concentration, and contralateral parotid superficial lobe cranio-caudal dose distribution) displayed agreement with our findings. Regressor importance changed when sub-cohorts were evaluated, but caudal aspects remained important. They concluded, however, that minimizing dose to the lateral and cranial aspects would reduce xerostomia incidence. Our relative importance assessments are in broad agreement, but our conclusions about clinical relevancy differ. Owing to the complexity of head-and-neck anatomy, minimizing dose to lateral and cranial aspects generally requires *increasing* caudal aspect dose. As we collectively have found caudal aspects to be important for clinical outcomes, the recommendation is surprising and implies our interpretation of *prediction* importances and *outcomes* importances differ. In recent work by Clark *et al.*, a model-based approach incorporating sensitivity analysis was used to assess relative importance (using the present cohort; see chapter 12). Linear models performed best and the collective caudal aspect slopes were both most important and largest in magnitude, implying that shifting dose to the caudal aspects would overall negatively impact salivary function. Similar findings have been reported by others [374]. We therefore believe that regressor importance (in this case) translates to clinical relevance. Differences in study designs, outcomes, assessment, cohort size and demographics, and factors considered (especially their response shape) may have contributed to

the discrepancy. However, our clinical recommendations are in agreement when the caudal aspects are dose-saturated and cranial or posterior aspects can be spared by shifting dose to the (already saturated) caudal aspects, which may reduce dysfunction. This common clinical situation demonstrates that characterization of regional effects throughout the entire parotid can improve outcomes risk analysis compared to simple recommendations to spare specific regions or lobes.

Ortholan *et al.* found in 2009 that the contralateral parotid volume receiving ≥40Gy (V_{40}) was the best dose-volume factor for predicting recovery of salivary function [449]. This finding suggests the non-equivalency of whole mean dose and V_{40} – both of which are dose-volume measures. Deviations from expected dose-volume effects, which follow directly from inhomogeneous radiosensitive structure distribution, have been known for several decades [353, 463]. While the findings of Ortholan *et al.* do not specifically describe a regional effect, the regions selected by our two approaches may overlap. Since standard clinical practice involves preventative irradiation of lymph nodes in the head-and-neck, proximate caudal parotid aspects often receive the highest dose. Therefore, V_{40} may simply be selecting the aspects, which would represent a dose-volume manifestation of a regional effect. We believe the reverse (caudal aspect importance reflecting V_{40}) is not true because contralateral parotid (lower dose) was found to be more important than ipsilateral parotid (higher dose), and axially the regions of highest dose follow a medial-anterior to lateral-posterior ridge. Both, along with the mean dose heat maps of fig. 13.4, demonstrate that importance does not merely reflect the dose profile. On the other hand, it is consistent that low-dose cranial-lateral sub-segments are the least clinically relevant and that sub-segments with greater relevancy are typically more heavily irradiated.

A more recent report by van Luijk *et al.* showed the presence of a confined critical region in the medial-dorsal aspects adjacent to mandible [357]. While our findings are not quite consistent in the superior-inferior direction, they appear to coincide in the anterior-posterior and medial-lateral directions. Both may coincide with major ducts, vasculature, or interfere with innervation; previous real-time imaging of stimulated parotids showed increased perfusion

variability focused in the vicinity of both regions [3]. The hypothesis that damage to stem/progenitor cells is the underlying cause of dysfunction, if true, would support ducts rather than vasculature or nerve impairment. The conclusion of a well-confined critical zone, however, was not confirmed in this work. We found that even very small regions are not necessarily 'critical.' At best, the most important sub-segments appear to have 4× the importance that a homogeneous parotid sub-segment would. It is possible that population averaging has 'smeared' importance. On the other hand, importance of the most caudal-anterior aspects were, in some cases, two orders of magnitude or greater than cranial and posterior sub-segments and naturally formed smooth importance gradients, which suggests an *effectively* critical (but somewhat broad and smeared) clinically relevant region. A smeared critical region would be more consistent with Lyman normal tissue complication probability models with parallel volume dependence parameters than confined critical regions, and may more accurately reflect stem/progenitor cell distribution [374]. Additional work is needed to characterize this effect.

Both Buettner *et al.* and van Luijk *et al.* report observing a bath-and-shower effect, which may confound importance assessment, especially for intensity-modulated radiotherapies. Knock-on effects (indeed, also higher-order interactions) are accounted for in RF and c-trees by permutation-based importances [458]. Explicitly including all first-order interaction terms for verification was not feasible even for segmentation into 18^{ths} owing to the increased complexity and decreased statistical power (i.e., a total of 171 regressors would need to be considered; n.b. N=332). Heterogeneous segmentation could in principle alleviate such issues, but it then becomes unclear how to robustly map regressor importance to clinical relevance.

Finally, the uncertainty in our findings were hard to directly quantify. The most reliable technique, c-trees, are memory-bound and computationally demanding. The bootstrap method is the most widely recommended method of quantifying uncertainty. Performing just 500 bootstraps via cloud computing would cost an additional $30-40k USD. Uncertainty quantification through bootstrapping is therefore not currently feasible. Instead, consistency of the derived importance maps, RF and c-trees methods, and importance

techniques were used to gauge uncertainty. All methods were in agreement that the caudal aspect is most important for salivary dysfunction. Furthermore, the relatively smooth importance gradients observed emerged naturally since neither RF nor c-tree methods had access to sub-segment localities. While it is not yet possible to directly quantify uncertainty in the importance maps generated by this work, reliability is thought to be high owing to the consistencies.

13.4.1 Conclusions

Caudal-anterior aspects of the parotid were found to be most important for prediction of radiation-induced late baseline-normalized salivary flow. Conditional inference trees, combined with fine segmentation, were found to significantly outperform whole parotid mean dose for prediction of salivary dysfunction.

Chapter 14

Other Regional Effects

14.1 Introduction

The non-parametric methods developed in chapter 13 are general-purpose methods that can be applied with little modification in other domains. In this chapter, the most robust and best-performing method (c-trees) is applied in two limited follow-up analyses. In section 14.2 regional effects are investigated in parotid using xerostomia questionnaire responses as response variables, and in section 14.3 regional effects are investigated in submandibular glands using resting saliva as the response variable. These two facets, xerostomia and submandibular glands, were chosen both to investigate congruence with the results of chapters 12 and 13 and because they are potentially clinically relevant in their own right.

14.2 Parotids and Xerostomia

Because each questionnaire response can be considered separately from all the others, analysis of xerostomia can produce a plethora of independent analyses and require careful control to mitigate the multiple comparison problem. Segmentation into 18^{ths} was chosen to reduce the computational load in this limited follow-up analysis of xerostomia-based regional effects.

Using only patients with complete xerostomia QoL questionnaires at

baseline, three months, and either one or two years, the most successful aspects of the non-parametric analysis of chapter 13 were applied using mean-scaled subjective responses. The same exclusion criteria (section 12.2.1) and mean-scaling technique (section 12.3.1) of chapter 12 were used. A total of 218 patients were found to be suitable for analysis.

QoL instrument questions are described in section 7.1. Each of questions 2-9 were assessed individually. The c-trees method was able to accurately predict questionnaire responses for each question; Pearson's correlation coefficients between actual vs. predicted responses for early and late xerostomia are shown in table 14.1. They varied from 0.618 to 0.674 which signifies uniformly 'strong' correlation (i.e., $r \in [0.60, 0.80)$, a threshold recommended by Evans [270]).

Question Number	r (Early)	r (Late)
2	0.645	0.647
3	0.664	0.624
4	0.655	0.640
5	0.674	0.648
6	0.642	0.636
7	0.645	0.654
8	0.633	0.654
9	0.618	0.674

Table 14.1: Pearson's correlation coefficients (r) between patient self-reported xerostomia questionnaire responses and the responses predicted using only mean dose to 18 equal-volume parotid gland sub-segments. Both early (i.e., three month) and late (i.e., one year and mean-scaled two year) responses were used. The QoL instrument questions are described in section 7.1.

As with parotids in chapter 13, heat maps were generated to visually display relative sub-segment c-tree conditional permutation importance (cf. fig. 13.3). Heat maps and relative importance ratios were overall similar with the caudal aspects being most important. Heat maps for questions[1] 6

[1] Question #6 was "*rate the dryness in your mouth while not eating or chewing*" and question #7 was "*rate the frequency in sipping liquids to aid in swallowing food.*" These two questions were selected randomly.

and 7 are shown in fig. 14.1. Quantitative relative importances are shown in figs. 14.2 and 14.3.

Figures 14.2 and 14.3 show importances that are *negative*, which demonstrate the increased uncertainty in this analysis compared to the parotid vs. stimulated saliva case, possibly due to fewer datum being available, or xerostomia responses being more variable and/or bimodal. Regardless, the same strong preference for caudal aspects of the parotid seen to be important for dysfunction are also found to be important for xerostomia. However, unlike the case of dysfunction posterior aspects appear to be more relevant than anterior aspects for xerostomia. To first order, both toxicities are in agreement that the caudal aspects are most important. Whether this is fortuitous or merely a consequence of the relationship between xerostomia and salivary dysfunction is not entirely clear. However, it suggests that sparing radiation dose to the caudal aspect may reduce *both* salivary dysfunction and xerostomia.

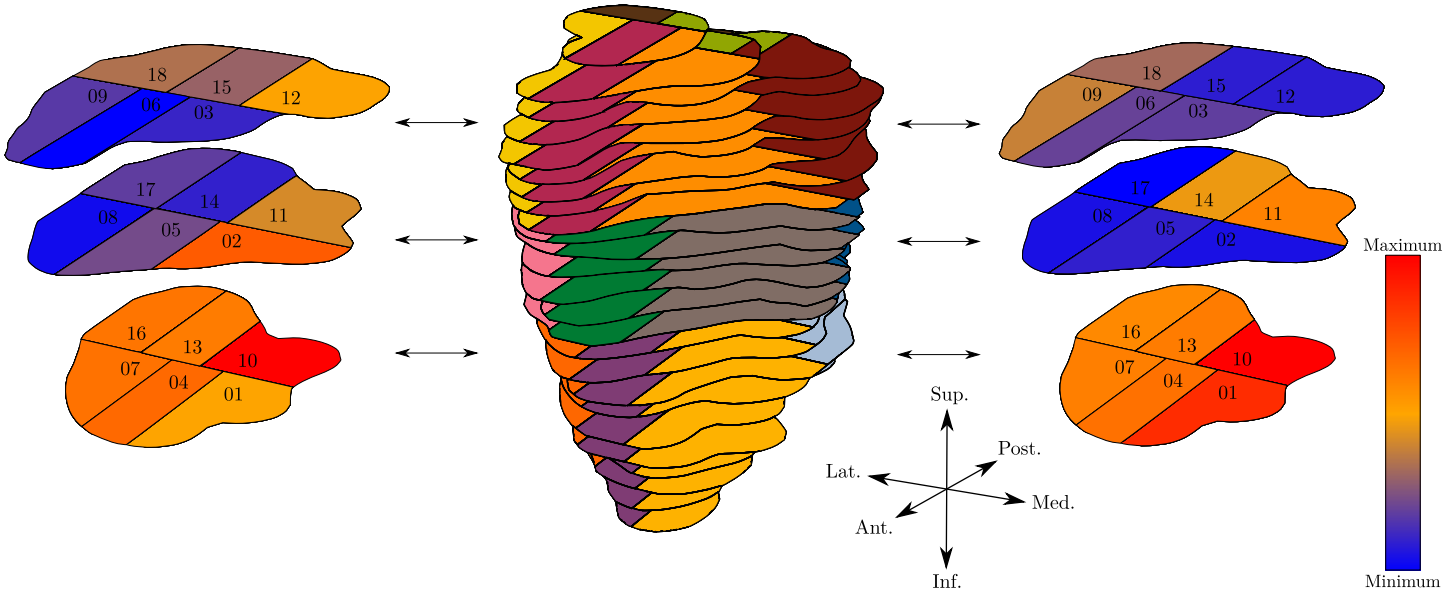


Figure 14.1: C-tree conditional permutation regional importance map using parotid gland sub-segment mean dose and patient self-reported late xerostomia questionnaire responses (#6 on left, #7 on right). Questions are described in section 7.1. Like with stimulated saliva, caudal aspects are most important (cf. fig. 13.3).

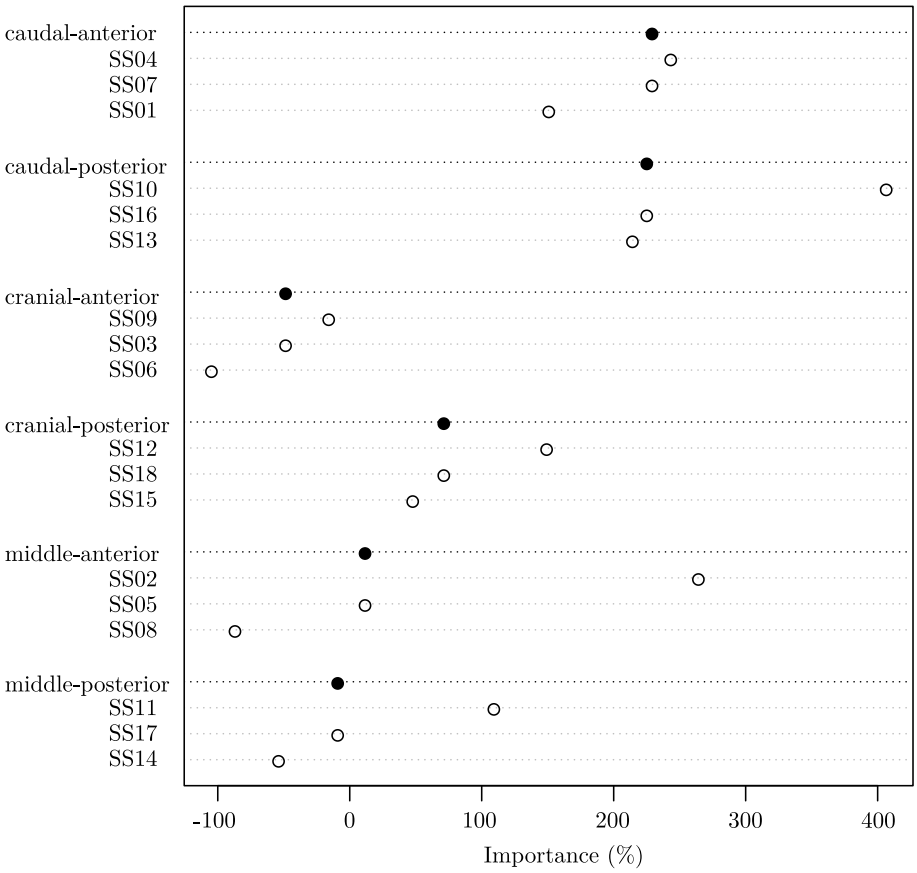


Figure 14.2: Quantified c-tree conditional permutation importance for individual sub-segments corresponding to QoL question #6. Sub-segment ('SS') numbering is shown in fig. 14.1 and is identical to fig. 13.3. Importance is presented as the percentage of relative importance compared to a homogeneous organ (which would be 100%). Anatomical groupings display the per-group median (filled circles). Caudal-posterior aspects are most important, with relative importance up to ∼4.0× that of an equivalent homogeneous organ sub-segment.

14.3 Submandibulars and Unstimulated Flow

Submandibulars are more oblique to the axial plane than parotids. While parotids have a characteristic inverted pyramid shape along the superior-

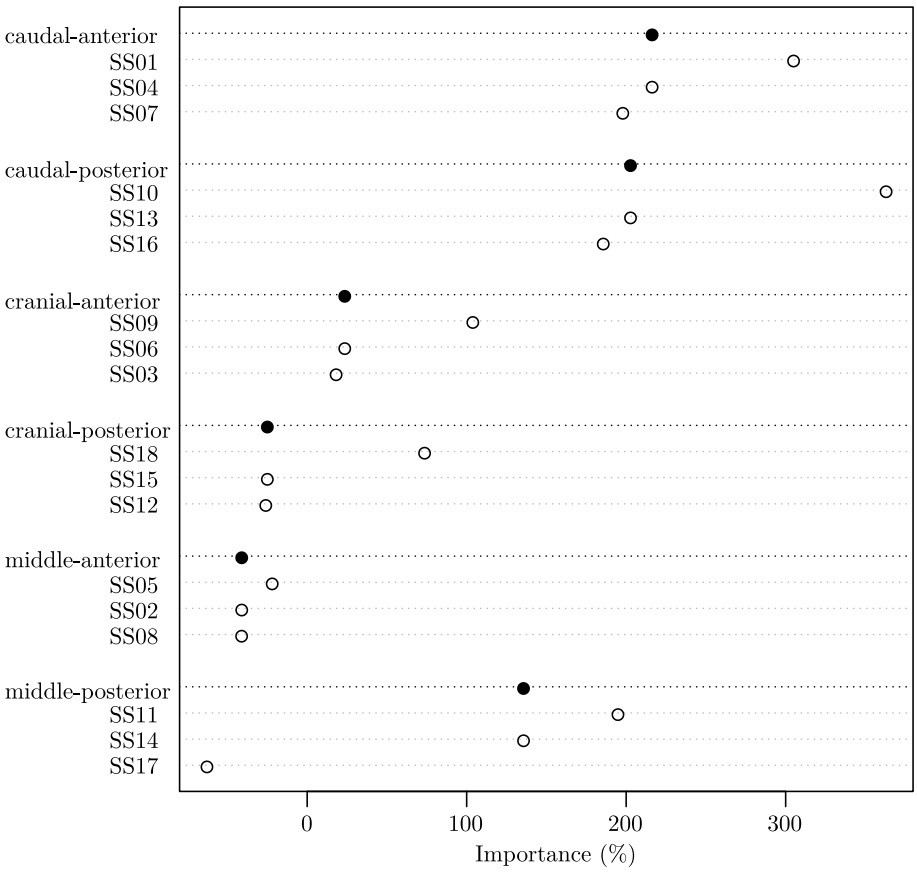


Figure 14.3: Quantified c-tree conditional permutation importance for individual sub-segments corresponding to QoL question #7. Sub-segment ('SS') numbering is shown in fig. 14.1 and is identical to fig. 13.3. Importance is presented as the percentage of relative importance compared to a homogeneous organ (which would be 100%). Anatomical groupings display the per-group median (filled circles). Caudal-posterior aspects are most important, with relative importance up to ∼3.5× that of an equivalent homogeneous organ sub-segment.

inferior axes, submandibulars are aligned more so along the anterior-posterior and superior-inferior axes. Segmentation is therefore slightly problematic. As submandibulars are demarcated on axial slices, there is a significant blurring of directionality. For example, if volumetric segmentation were used to divide a submandibular into two halves using an axial plane, the resulting superior-most sub-segment may include substantial inferior portions due to the obliquity.

There are several ways to overcome this problem. The simplest way would be to find the longest line-segment that can be fully enclosed within the ROI. This method suffers from topological sensitivity, but will locate the longest dimension of the submandibular to first order. A more robust method would employ Principal Component Analysis (PCA), which is conceptually similar in purpose to the simple method, but can be used to generate a complete orthogonal set of directions along which planar segmentation could be aligned. PCA is also more robust to topology as it can be made to take into account volume (or density).

However, the issue of *whether it is worthwhile* to correct this problem is not clear. While it would give a planar segmentation that is independent of patient positioning (though not necessarily incorporating deformation due to positioning), it would also present an additional barrier to clinical applicability. In particular, if segmentation is performed along axes respecting the existing (axial) orientation, then it will be considerably easier clinically to determine sub-segment mean doses. Furthermore, it is not clear *a priori* that submandibular morphology, like parotid morphology, holds any importance for regional effects. The distribution of functional sub-units may not respect topology or morphology anyways. For these reasons, PCA was not employed.

This analysis procedure used the same mean-scaling approach and exclusion criteria as described in chapter 13, except that surgeries that interfered with submandibulars rather than parotids led to patient exclusion. As seen in table 3.2, fewer submandibulars are contoured compared to parotids due to their increased clinical relevance. In total, 314 patients were found to be suitable for analysis.

Segmentation into 8^{ths} was performed to ease analysis but still provide

minimal anatomical groupings along the three cardinal axes. A heatmap for c-tree permutation importance is shown in fig. 14.4 and a heat map for c-tree conditional importance is shown in fig. 14.5. A quantitative importance chart that shows differentiation between cranial and caudal aspects is shown in fig. 14.6.

Similar to parotid, the submandibular demonstrates a clear regional effect. The most important sub-segment (SS04) is the cranial-posterior sub-segment nearest to the floor of the mouth. Relative importance is $\sim 3.5\times$ that of an equivalent homogeneous organ sub-segment. However, like section 14.2, the link between relative importance and clinical relevance was not investigated and is only assumed to be linear in the sense of a first-order approximation.

It is curious that the regional effect seen in submandibulars is transversely inverted compared to parotids. It is not possible to identify if this result is merely coincidental or the product of some underlying phenomena. However, speculatively, it is possible the apparent inversion has an anatomical basis, or a knock-on effect between parotid and submandibular glands exist such that high dose to both simultaneously produces a greater effect than would be had by merely superimposing the effect of each individually, or that dosimetric coupling between both glands creates a 'pocket' of undersampled data (i.e., high parotid gland dose with low submandibular dose, and vice-versa). While the parotid gland-only analysis did not suffer this ambiguity (since the ipsi- and contralateral glands are usually well differentiated), a more sophisticated analysis combining multiple facets is needed to provide a basis for inter-organ importance comparisons.

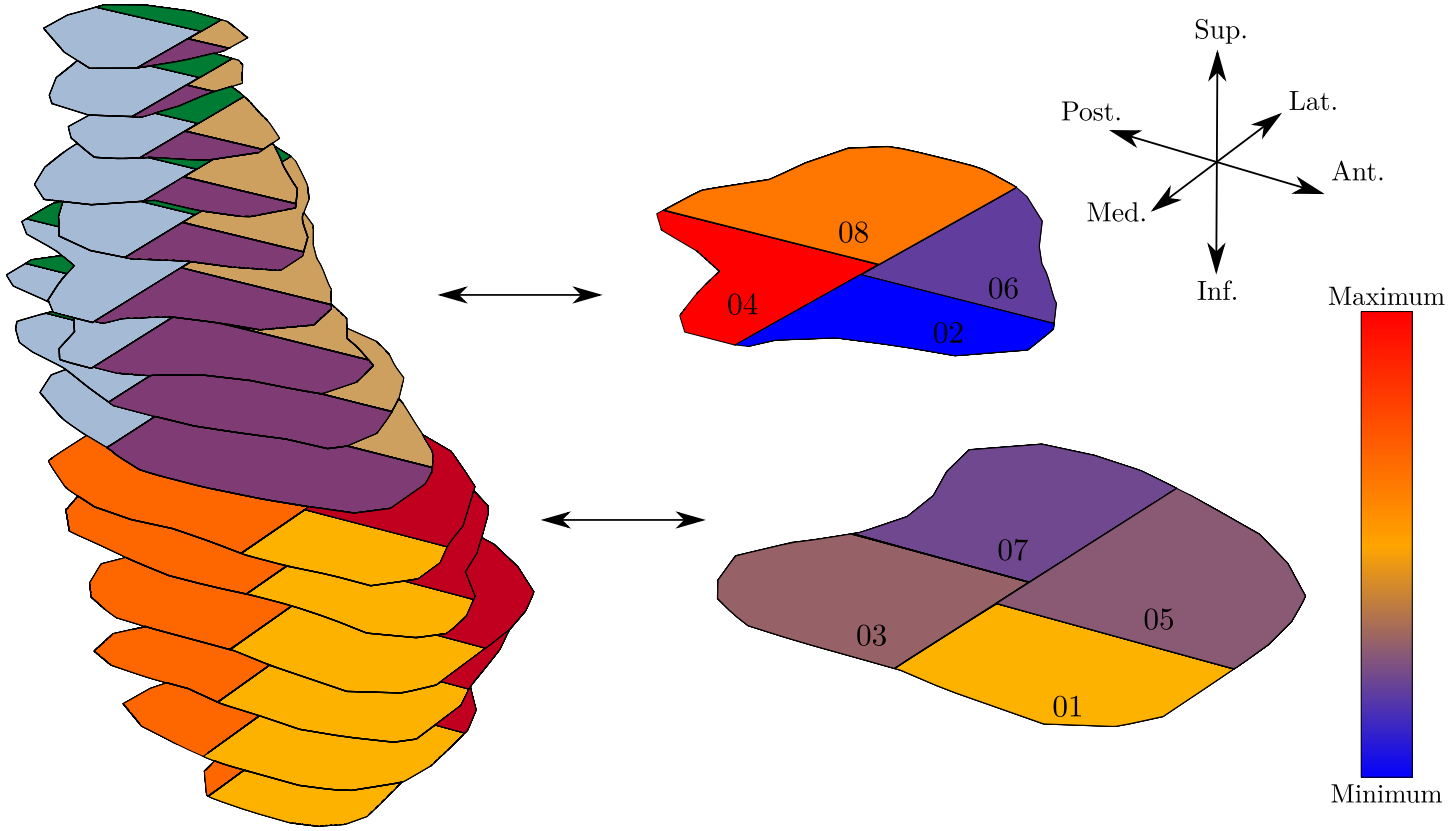


Figure 14.4: C-tree permutation (non-conditional) regional importance map using submandibular gland sub-segment mean dose and late resting saliva facets. Cranial aspects (closest to the floor of the mouth) are most important.

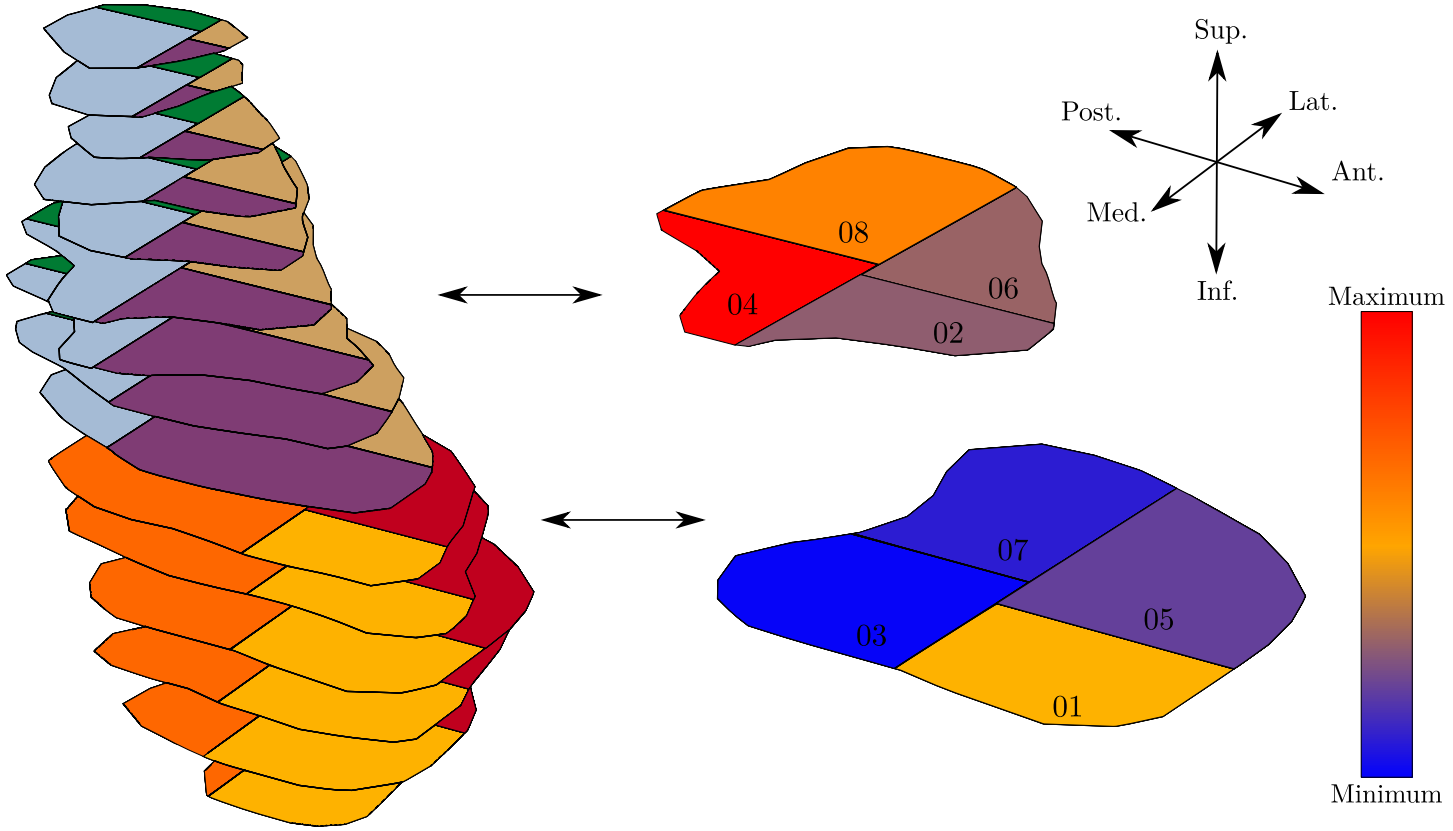


Figure 14.5: C-tree conditional permutation regional importance map using submandibular gland sub-segment mean dose and resting saliva facets. Cranial aspects (closest to the floor of the mouth) are most important. Agreement with c-tree non-conditional importance (in fig. 14.4) is strong.

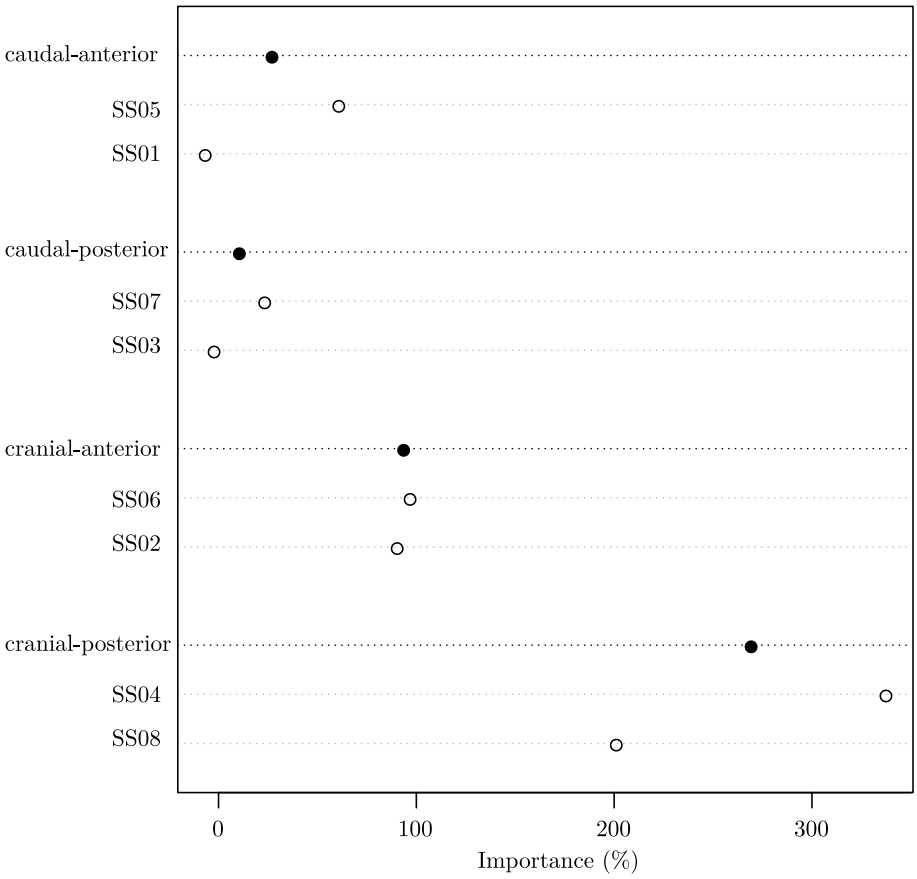


Figure 14.6: Quantified c-tree conditional permutation importance for individual sub-segments. Sub-segment ('SS') numbering is shown in figs. 14.4 and 14.5. Importance is presented as the percentage of relative importance compared to a homogeneous organ (which would be 100%). Anatomical groupings display the per-group median (filled circles). Cranial-posterior aspects (closest to the floor of the mouth) are most important, with relative importance up to $\sim$3.5$\times$ that of an equivalent homogeneous organ sub-segment.

14.4 Summary and Conclusions

In this limited follow-up analysis, the most consistent methods of chapter 13 were applied to analyze resting saliva, submandibulars, and xerostomia. Regional effects were found. The caudal aspect of the parotid, like the stimulated saliva case, were most important for xerostomia. However, anterior and posterior aspects were less differentiated. The submandibular showed different regional importance, with the cranial aspect being most important.

The meaning of importance in chapters 12 and 13 was derived from a combination of model selection, sensitivity analysis, and permutation techniques. Linear models, which were the best candidate models in every case considered, established a link between importance derived from analysis and clinical relevance. However, this association is not fully understood for submandibulars or xerostomia. In both cases, a more careful inspection is required to assess the relationship. The purpose of this chapter was not to fully demonstrate regional importance, but rather to motivate further research on the topic.

Chapter 15

Development of a DCE-MRI Imaging Protocol[1]

15.1 Introduction

Saliva supports ongoing tooth and gum solidity in various ways: by acting as a buffer against acids and bases; maintaining a healthy oral flora by flushing bacteria and mastication debris from the oral cavity; delivering digestive enzymes to the oral cavity; and easing oral transport by lubricating oral surfaces. These basic functions enable a variety of everyday facilities, such as speech, efficient mastication and deglutition, and the perception of taste. Consequently, the loss of salivary function has a strong impact on one's quality of life [464]. Salivary function loss resulting from radiotherapy treatment of head-and-neck cancers is common [236].

There are many critical tissues in the head-and-neck (e.g., spinal cord, brainstem, ocular nerves) for which delivering even moderate radiotherapy doses results in catastrophic repercussions for patients. On the other hand,

[1]The contents of this chapter are an updated version of an early manuscript that was later published under the title '*Development of a method for functional aspect identification in parotid using dynamic contrast-enhanced magnetic resonance imaging and concurrent stimulation*' in Acta Oncologica (2015 Oct 21; vol. 54, no. 9, pp. 1686-90; Taylor & Francis Ltd., http://www.tandfonline.com/loi/ionc20) by Haley Clark, Vitali Moiseenko, Thomas Rackley, Steven Thomas, Jonn Wu, and Stefan Reinsberg [3].

delivering the prescribed dose to a suspected tumour volume is crucial for maintaining local control. Spectator tissues, which are not critically radiosensitive nor part of the tumour volume, are used as conduits for delivering radiotherapy dose to inaccessible tumour volumes. Highly radiosensitive tissues are spared by shifting dose to spectator tissues like parotid glands and other salivary organs [465]. Though such irradiation is unavoidable, treatment planners using modern treatment techniques like volumetric modulated arc therapy or intensity modulated proton therapy have freedom to adjust intra-parotid radiation quantities and locations. Detailed knowledge of tissue response and outcome risk is therefore needed for effectual planning [466, 467].

Presently used consensus guidelines for parotid gland sparing assume a homogeneous distribution of functional burden [236], are difficult to attain in practice, and do not ensure specific outcomes [410]. In recent years, evidence has mounted to suggest a heterogeneous distribution of functional burden within the parotid [466]. Recent reports have found delivering dose to one region of rat parotid results in a different incidence of objective xerostomia (dry mouth) than delivering that same dose elsewhere [356]. Regional susceptibility of subjective (i.e., patient-reported) xerostomia has been noted in human parotid [233]. In light of an earlier investigation by our group, it is presently unclear to what extent these findings relate to objective function alteration in humans, or whether such regions align with parotid parenchyma [324]. Pursuit of this avenue of research is enticing owing to the potential ramifications on treatment planning and possible improvement of outcomes for head-and-neck patients.

In a recent survey of salivary gland radiation reduction techniques, Vissink *et al.*. [465] advocate tissue sparing as the most effective method. There is growing evidence that functional imaging can be clinically relevant for more clearly defining target volumes and assessing adverse normal tissue effects [466]. We report the development of a novel technique making use of Dynamic Contrast-Enhanced Magnetic Resonance Imaging and concurrent salivary stimulation which can potentially identify parotid parenchyma in healthy volunteers. Inter- and intra-parotid tissue differentiation are possible, and

application of the protocol could potentially improve tissue sparing. Results from a small, healthy volunteer trial are provided.

15.2 Methods

15.2.1 Ethics and Accrual of Volunteers

The study protocol was approved by the BCCA University of British Columbia (UBC) REB, and is in accordance with agency ethical standards and the Helsinki Declaration. Between December 2014 and May 2015, four healthy individuals (one female and three males, between 25 and 35 years of age) volunteered for this study, giving informed consent. Individuals were excluded if they presented any standard MR contraindications (e.g., incompatible implants, prosthetics, or clips; foreign metallic bodies, including shrapnel or debris in their eyes; pregnancy), contraindications to intravenous gadolinium injection (e.g., history of adverse reactions, history of or family history of renal disease), or had metal retainers or amalgam fillings which could cause susceptibility artifacts. For the purposes of this study symmetrically paired organs (parotids, masseters) are effectively treated as individual organs, resulting in a total pool of eight unique organs.

15.2.2 Image Collection and Processing

Perfusion imaging was chosen as the primary imaging method due to its power to non-invasively characterize functioning glandular tissues with high temporal resolution. Reports have demonstrated the ability of Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) to quantitatively measure radiotherapy-induced parotid gland perfusion alterations. These alterations are thought to result from increased extracellular-extravascular space and decreased vascular permeability stemming from radiation damage [468]. Likewise, function alteration can also be quantified via blood perfusion to parenchyma [469]. Perfusion changes measured via DCE-MRI reflect physiological changes.

Magnetic resonance images were collected using a dedicated research 3T

Philips Achieva MR scanner at the University of British Columbia. The imaging protocol is as follows. Volunteer hydration prior to imaging was controlled by asking them to abstain from food or drink the night prior and leading up to the imaging session. DCE images used intravenously administered gadolinium contrast agent (Bayer Magnevist gadopentetate dimeglumine) at the manufacturer's recommended total dose of 0.2mL/kg with an injection rate of 4mL/s followed by a 20mL saline flush. DCE imaging comprised a T1-weighted spoiled gradient echo sequence with 2.960ms repetition time, 1.351ms echo time, and a flip angle of 12°. Images were reconstructed axially with $1.46 \times 1.46 \times 4.00$mm spatial resolution, 3.9s temporal resolution, 375mm field of view in both phase-encode and frequency-encode directions, and 4mm slice thickness. Images were continuously collected for 450-600s. Contrast agent was administered 45-60s after DCE sequence commencement. Salivation was manually induced 170-240s after contrast agent injection by passing a small amount (8mL) of a weak citric acid solution (2% by weight) into the oral cavity using a syringe via polyethylene tubing.

Additional short gradient echo scans with a 3° flip angle were collected prior to contrast injection. These were compared to the aforementioned longer-running 12° flip angle images to generate per-voxel contrast agent time courses C(t). To expedite total scan time by forgoing collection of images with additional flip angles, the signal difference method discussed by Ashton [470] was used to compute C(t). Specifically, pre-contrast signal was averaged and subtracted from post-contrast signal. Spatial averaging was used to reduce the impact of noise.

For three of the four volunteers, two DCE sequences were performed back-to-back. In the first, 1/3 of the total contrast agent was injected, no stimulation was performed, and 450s of data was continuously collected. In the second, the remaining 2/3 was injected, salivation was stimulated, and imaging continued for 600s. Injection splitting was done to produce a baseline contrast agent response curve from which we could more clearly access the stimulatory response. Scans without stimulation were shortened to reduce total imaging duration. Splitting 1/3 – 2/3 produced a baseline curve without substantially reducing the second scan signal-to-noise ratio with lingering

contrast agent.

ROIs (parotid, masseter, and pharyngeal tissues) were manually contoured from anatomical and DCE MR images using the DICOMautomaton software suite [7]. Parotid ROIs were also partitioned into anterior/posterior halves.

15.2.3 Statistics – Variance Analysis

Non-parametric techniques were employed. The primary technique developed for inter-parotid analysis (and cross-organ analysis, e.g., parotid vs. masseter) is a topological analysis and involves characterizing the variance of C(t) on a cluster-of-voxels basis. We refer to it hereafter simply as 'variance analysis.' Variance was selected because the observed time courses proved difficult to model but were uniformly more variable after salivation was induced. The procedure is straightforward[2]. First, for each voxel in a given ROI, at a specific temporal point, the contrast agent present was computed by averaging neighbouring in-plane voxels. For the entirety of this report, voxels were included if they were less than two voxel-widths (2×1.46mm) away from the centre of a given voxel. The specific number of neighbours used did not appreciably affect findings. Second, time courses (spatially-averaged contrast over time) were constructed for each voxel. Third, variance was estimated using an unbiased estimator over a temporal sliding window. For the entirety of this report, the window extended $\pm$20s from a given datum. While conclusions were not substantially affected by the width of the window, spikes in the resulting time courses broaden as this width is increased. Fourth, variance time courses were combined over the ROI by summation. Finally, because summed variance time courses depend on ROI volume, the total amount of contrast injected, and other factors, they were normalized to one another in the window of time after the initial contrast agent uptake peak but before stimulation occurred.

Variance analysis admits a natural way to compare pairs of time courses: compute the difference between the curves, compute the mean of the difference over all time points before and after stimulation, and examine whether the

[2]Not only is it straightforward computationally, requiring only basic descriptive statistics, but it also has no underlying model which can be violated.

means are significantly different. The same procedure is applicable without subtracting curves; instead of comparing the mean of the differences, the mean absolute variances can be compared. Since the null hypothesis is that there is no response to stimulation, and variance should remain approximately constant during the washout period, the mean will not change unless there is a stimulatory response.

Means are compared using the standard two-sample unpooled t-test for unequal variances (Welch's t-test). Approximate normality of differences was explicitly inspected in all cases. The paired Wilcoxon signed-rank test is used to assess whether population mean ranks differ between two paired samples. Both tests are invariant to uniform ordinate scaling, so for simplicity arbitrary units are used throughout.

15.2.4 Image Maps

A qualitative, non-parametric, topological method was developed to characterize intra-parotid tissue variation before and after stimulation. The technique is as follows. First, spatially averaged contrast agent time courses are constructed (as above) for each voxel in both stimulated and unstimulated DCE series. Second, each series is broken into pieces and two parts are retained: (1) post-contrast agent injection and pre-stimulation, and (2) post-stimulation. For unstimulated series, the stimulation break point was taken to be identical to the stimulated series break point. Third, a straight line was individually fit to each part (before and after stimulation). Fourth, the difference of slopes in the unstimulated case were subtracted from the difference of slopes in the stimulated case. As the shape (in contrast to the scale) of C(t) is not strongly affected by the specific quantity of agent, and is monotonically decreasing in the washout phase, any residual quantity indicates a response (i.e., variations in washout) to stimulation. Finally, a map is generated from the residual quantity. This technique is later referred to as the "difference of changes in slope" technique.

15.3 Results

The mean parotid volume was $19.9 \pm 4.4\text{cm}^3$ (mean $\pm\sigma$ of the mean, median: 18.2cm^3). Left and right volumes correlate strongly (mean for left: 19.6cm^3 vs.right: 20.2cm^3).

A typical, spatially averaged per-voxel C(t) from parotid is shown in fig. 15.1. Key features of the protocol are visible. From left to right: (1) the pre-contrast agent injection window (left-most grey box); (2) the rapid uptake period, where high concentrations of contrast rapidly perfuse into the parotid tissues, peak, and then begin to slowly wash out; (3) the period of stimulation – in this case, beginning at 230s and continuing until approximately 240s from the scan start – and a response to stimulation manifest as a modest increase in contrast agent concentration; and (4) the continued slow contrast washout.

Not all C(t) look as clear as fig. 15.1; the volunteer in fig. 15.1 received a single injection of the full contrast agent to maximize the signal-to-noise ratio. Examples of other C(t) are shown in fig. 15.2. In particular, some voxels decrease, rather than increase, and many do not show any obvious response whatsoever. Response, if any, is generally delayed 10-30s after stimulation commencement. The response shape varies from a fast positive or negative 'blip' to an ongoing shift or bias.

15.3.1 Variance Analysis

For 3 of 4 volunteers (6 of 8 parotids), variance analysis demonstrated a clear distinction in apparent parotid stimulatory response (see fig. 15.3). The fourth volunteer's variance was uniformly high and consequently no response was detected. The mean of the difference before and after stimulation in fig. 15.3 were significantly different (two-tailed t-test: $0.0 \pm 1.1E-4$ pre- vs. $1.9 \pm 0.1E-3$ (arb. units) post-stimulation mean $\pm\sigma$ of the mean; $p < 0.0001$). It is visually apparent that the paired Wilcoxon sign-rank test was also significant as the curves are almost entirely separated by a large gap post-stimulation ($p < 0.0001$).

In 4 of 6 parotids (2 of 3 volunteers) where both stimulated and unstimulated time courses were collected, the Wilcoxon test indicated a significant

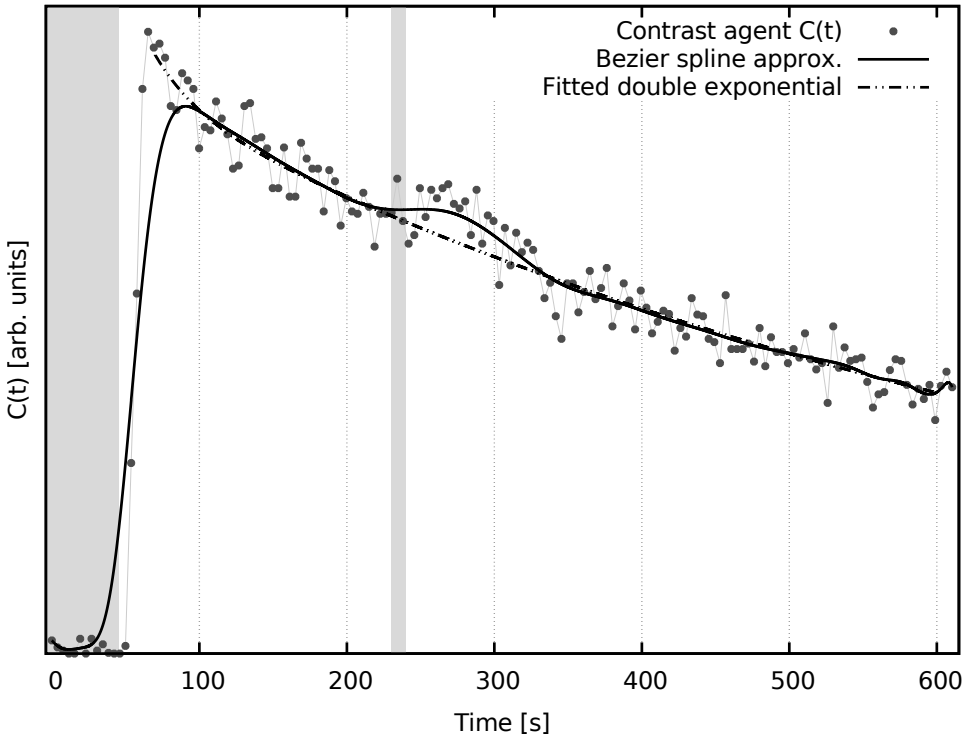


Figure 15.1: A typical spatially averaged voxel C(t) demonstrating temporal stages of the protocol. From left: pre-contrast agent injection window (left-most grey box); rapid uptake period, where high concentrations of contrast rapidly perfuse into parotid tissues, peak, and begin to drain; stimulatory period running from 230-240s from scan commencement, and a stimulatory response manifest as a modest contrast agent concentration increase; and continued slow washout. An empirical fit omitting the stimulatory period and Bezier spline interpolation are shown as visual guides. Figure previously published in [3], reproduced with permission from Taylor & Francis Ltd.; http://www.tandfonline.com/loi/ionc20.

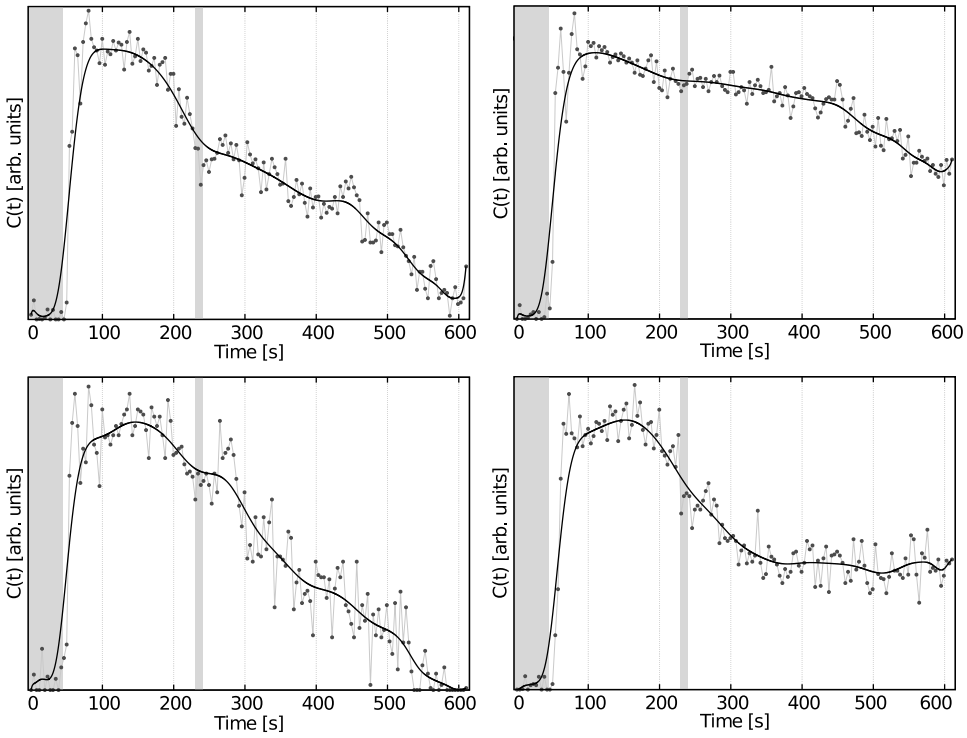


Figure 15.2: Examples of time courses similar to fig. 15.1, but showing varying responses to the stimulation beginning at 230s. Positive, negative 'blips' and ongoing shifts are seen. Splines are used as a visual guide; note the strong deviation 10-30s after stimulation. Figure previously published in [3], reproduced with permission from Taylor & Francis Ltd.; http://www.tandfonline.com/loi/ionc20.

difference in pre- and post-stimulation variances ($p < 0.05$). The other two parotids were near significance ($p = 0.05$ and $p = 0.08$). Two-tailed t-tests could be applied to examine mean shift (i.e., mean variance pre- vs. post-stimulation) in all cases; examination of the stimulated time course showed that 6 of 8 parotids had $p < 0.02$ ($1.88 \pm 0.17E-2$ pre- vs. $1.68 \pm 0.33E-2$ post-stimulation or greater separation). For unstimulated time courses, the mean shift for 3 of 6 parotids had $p < 0.05$.

For each individual volunteer, parotid (left and right) was significantly

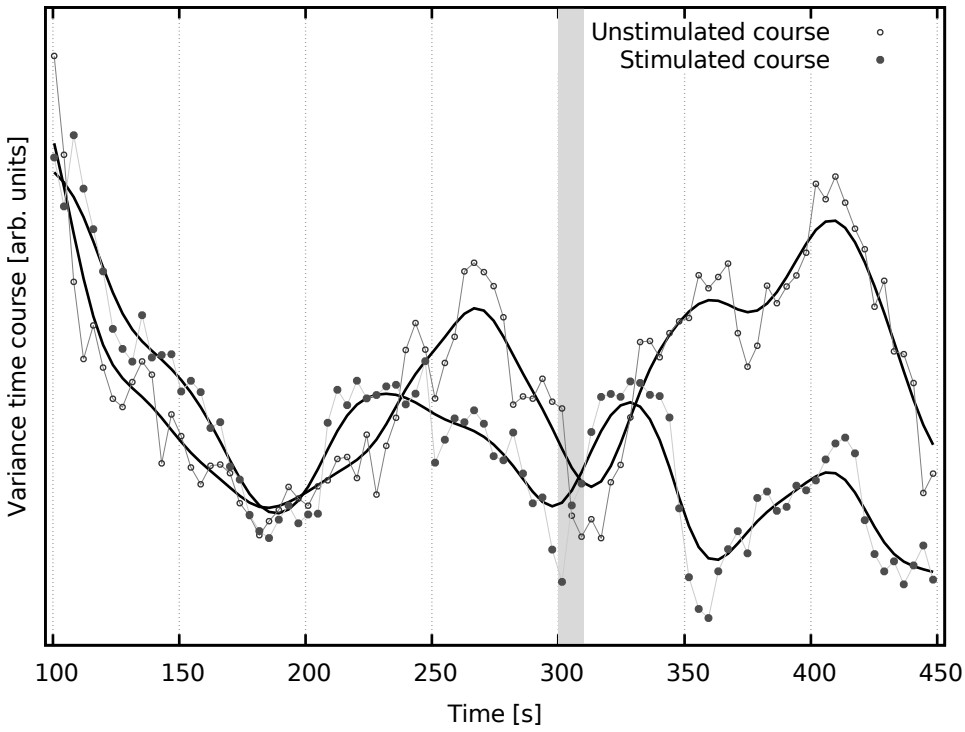


Figure 15.3: Variance analysis time courses in parotid (stimulated and unstimulated, with Gaussian kernel smoothed trend lines as a visual guide) showing a clear distinction in trend after stimulation occurs (300s). Means before and after stimulation are significantly different ($p < 0.0001$; two-tailed t-test), suggesting a differing contrast dynamics resulting from stimulation. Figure previously published in [3], reproduced with permission from Taylor & Francis Ltd.; http://www.tandfonline.com/loi/ionc20.

distinct from masseter (left and right) after stimulation in 13 of 16 cases: using a paired Wilcoxon sign-rank test, $p < 0.0001$ for 12 of the 13 and $p = 0.04$ in the remaining case. In 12 of 16 cases the mean of the differences were significantly different before and after stimulation using a two-tailed t-test ($p < 0.02$). Comparison of left and right masseter for each patient showed that in 3 of 4 cases, masseters did not respond differently to stimulation (two-tailed t-test $p > 0.07$; similar Wilcoxon p-values). Comparison of masseter to a variety of nearby, non-specific pharyngeal tissues in a single volunteer

indicated they were, on average, not significantly different (Wilcoxon $p = 0.07$; two-tailed t-test $p = 0.72$). However, comparison of left and right parotid show that they respond differently to stimulation: $p < 0.0001$ for all Wilcoxon tests, $p < 0.005$ ($0.9 \pm 4.2E-4$ left vs. $4.2 \pm 0.6E3$ right or greater separation) for all t-tests.

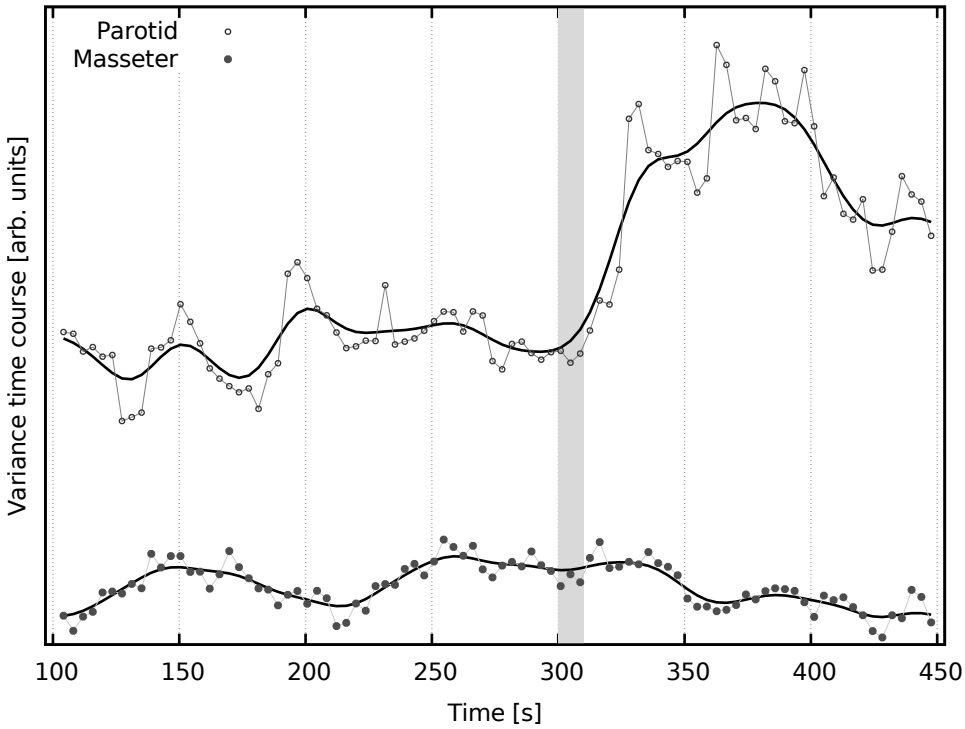


Figure 15.4: Comparison of parotid and masseter response to stimulation. Compared with nearby tissues, parotid response is more rapid and greater in amplitude. Figure previously published in [3], reproduced with permission from Taylor & Francis Ltd.; http://www.tandfonline.com/loi/ionc20.

The variance analysis technique found post-stimulation parotid to be significantly distinct from masseter, pharyngeal tissues, and other parotids. An example comparison is shown in fig. 15.4. A variance analysis was also run on posterior and anterior parotid portions of equal volume. Similar to whole parotids, two-tailed t-tests quantified mean shifts. For anterior parotid,

stimulated time courses in 6 of 8 parotids had $p < 0.05$ ($1.0 \pm 0.3E-2$ pre- vs. $9.7 \pm 0.7E-2$ post-stimulation), whereas for unstimulated only 2/6 had $p < 0.05$. For posterior parotid, stimulated time courses in 4 of 8 parotids had $p < 0.05$; 4 of 6 had $p < 0.05$ for unstimulated courses. Comparison of the anterior and posterior portions directly showed a significant discrepancy in stimulatory response in right parotid (Wilcoxon p 0.001) in 3 of 4 cases. The same discrepancy was seen in the left parotid in all 4 cases. These findings indicate that anterior and posterior aspects of the parotid show independently distinct responses to stimulation.

15.3.2 Image Maps

To further assess intra-parotid variations, image maps were generated using the difference of changes in slope technique. In these maps, a voxel that has no response to stimulation will be midtone. Voxels that respond with a positive change in slope are brighter, while those that respond negatively are darker. Example slices from two volunteers are shown in fig. 15.5 which clearly shows intra-parotid variation. An enlarged example is shown in fig. 15.6.

15.3.3 Discussion

The aim of this pilot study was to develop a DCE-MRI imaging protocol capable of identifying parotid gland parenchyma in healthy volunteers. No existing literature on concurrent DCE-MRI and salivary stimulation was found. DCE-MRI was chosen for its temporal resolution and ability to assess functional alterations via blood perfusion to parenchyma [469]. Scintigraphy is a well known and historically well used technique for quantifying parotid function but produces 2D images and requires the use of costly radioisotopes (see [464] and references therein). A novel technique making use of dynamic ^{11}C-methionine PET analogous to DCE-MRI has been described by Buus et al. [69, 471] which improves on earlier single photon emission CT methods in spatial resolution. PET produces high quality volumetric images but requires injection of a positron-emitting tracer and may require an additional imaging modality for (co-)registration [472]. Both potentially increase patient dose.

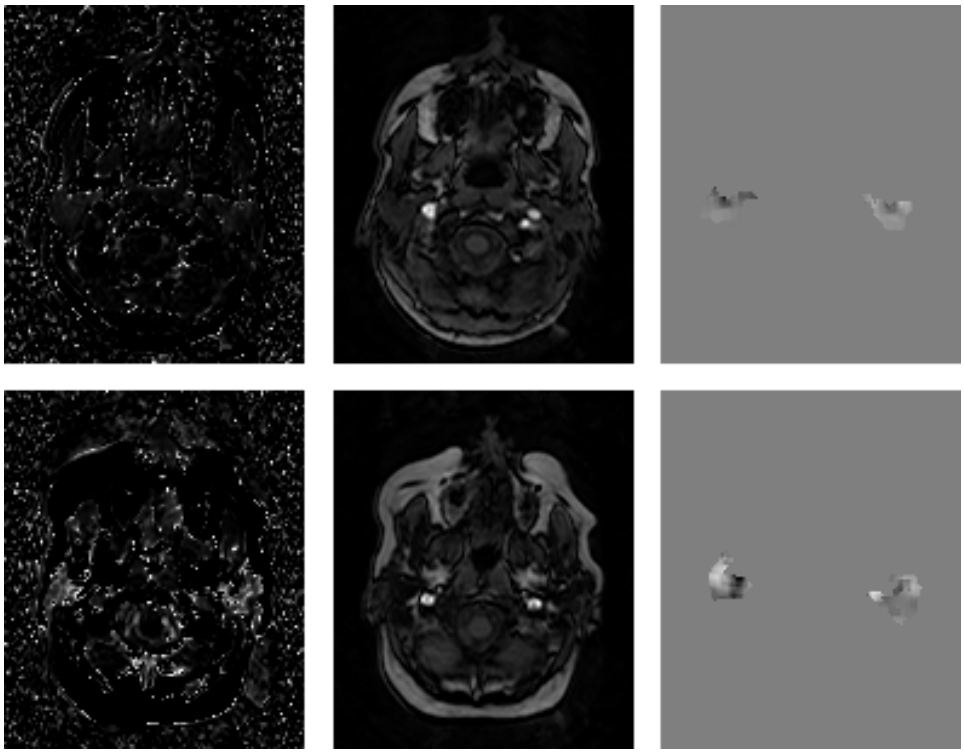


Figure 15.5: A single slice example of image maps for two volunteers (top and bottom). At centre: temporally-averaged T1-weighted images; at left: contrast agent; at right: difference of changes in slope maps in parotid. In the latter, voxels which showed no response to stimulation (within the ROI) are midtone, those that responded with a positive change in slope are brighter, and those that responded negatively are darker.

Perfusion computed tomography is generally considered a low-cost, viable alternative to DCE-MRI [81, 473], but DCE-MRI generally has superior spatial and temporal resolution, and requires no ionizing radiation [474].

Perfusive changes were observed following stimulation, but response varied. Figure 15.1 shows a typical spatially averaged C(t) from a parotid voxel. Examples of other C(t) are shown in fig. 15.2. Response, if any, was generally delayed 10-30s after stimulation commencement.

For 3 of 4 volunteers (6 of 8 parotids), variance analysis demonstrated a

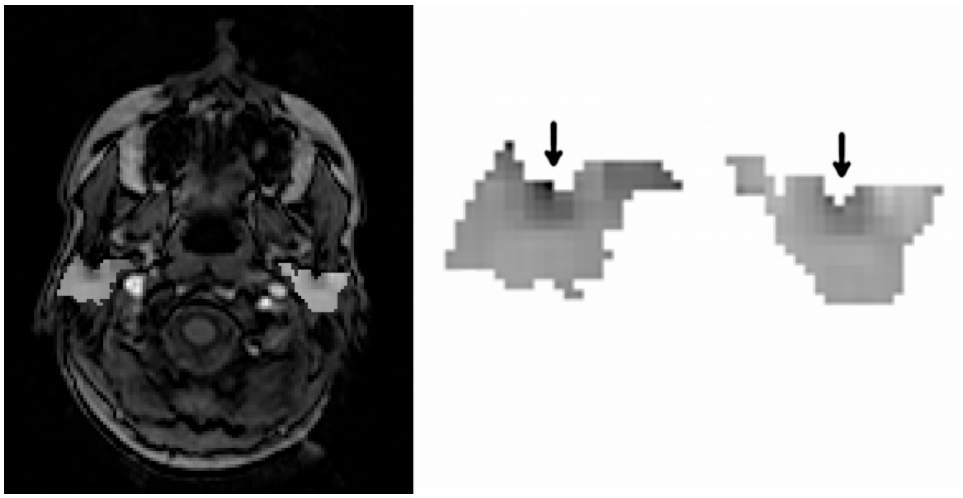


Figure 15.6: Enlarged example image map slice. At left: a temporally-averaged T1-weighted image with the difference of changes in slope map overlaid on the parotids; at right: enlarged parotid maps. In the latter, voxels showing no stimulatory response are midtone. Those that responded positively (negatively) are brighter (darker). Figure previously published in [3], reproduced with permission from Taylor & Francis Ltd.; http://www.tandfonline.com/loi/ionc20.

clear distinction in parotid stimulatory response. This result, combined with observed differences before/after stimulation and differences in pre-/post-stimulation variances, suggests variation in parotid response depending on the presence of a salivary stimulus.

Image maps figs. 15.5 and 15.6 were generated using the difference of changes in slope technique to assess intra-parotid variations. Such variation was observed. The portion of parotid nearest to the posterior edge of the mandible (as indicated) was most dissimilar from surrounding parotid tissues. This region was recently found by van Luijk *et al.* to house stem/progenitor cells in rat parotid, and was reported as being strongly correlated with post-radiotherapy salivary output in humans (personal communication[3], 2014). The alignment of regions found using different techniques suggests the

[3]Note: was ultimately published as [357].

proposed techniques may be suitable for locating critical regions. Maps were slowly varying and qualitatively regular across patients, suggesting a possible heterogeneous functional burden distribution.

Similar to the technique described by Buus et al. [69, 471], our variance analysis and difference of changes in slope techniques could be used to assess radiotherapy induced functional alterations. Unlike Buus *et al.*'s technique, through the use of MR, our technique could be used to directly observe regional salivary compensation in nearly real-time throughout the entire 3D ROI.

For each individual volunteer, parotid (left and right) was generally significantly distinct from masseter (left and right) after stimulation. Masseter was not distinct from nearby non-specific pharyngeal tissues. Left and right masseters did not respond differently to stimulation, but left and right parotid did. This indicates parotids are more strongly responding to stimulation than masseter. Distinction in parotid response was apparent. Variance analysis showed that after stimulation, parotid was significantly distinct from masseter, pharyngeal tissues, and other parotids. Compared with parotids, the response of nearby tissue to stimulation occurred later and with reduced amplitude. An example is shown in fig. 15.4. This finding is logical: salivation involves the transport of water which is rapidly replenished from the blood plasma during continued flow [25].

A variance analysis performed on posterior and anterior portions of the parotid showed that, on average, there was a significant discrepancy in left and right parotid stimulatory response, indicating that anterior and posterior parotid aspects show distinctly different responses to stimulation.

There are a number of limitations that remain to be addressed. It is not clear whether the proposed techniques can handle the so-called bath and shower effect observed in rat parotid [356], which complicates tissue sparing. An explanation for this effect proposed in [356] and observed by Konings et al. [348, 349] – that portions of the parotid can be regenerated by progenitor cells in distant portions – would require a more sophisticated analysis if correct. However, parenchyma localization and sparing would likely remain valuable for reducing early functional loss.

One limitation of the signal difference method is that it ignores non-linearity in the conversion from signal intensity to concentration that arise from simultaneous alteration of tissue T1 by the contrast agent [475]. Comparison between C(t) reconstructed using a traditional method via three flip angles and the signal difference technique showed the latter to be more stable. Flip angle variability was not explicitly measured and remains unknown. The signal difference technique is thought to be more reliable in this regime. A limitation of the difference of changes in slope technique is that slope changes represent a complicated admixture of pharmacokinetic parameters which cannot be easily interpreted as a specific change in tissues. Given that function alteration can be quantified via blood perfusion [469], perfusive changes are likely to play a strong role. Further investigation is needed.

Functional tissue localization could potentially be improved using more advanced, faster imaging techniques or supplementary imaging. Candidates include relaxometry [476], blood oxygenation level dependent MR [477], and intravoxel incoherent echo-planar motion imaging [80, 478]. De Langen *et al.* [472] suggests that dynamic PET and DCE-MRI are largely complementary techniques for assessing tumour blood flow – we believe simultaneous use of Buus *et al.*'s [69, 471] PET technique is also amenable to our method and may enable functional structures to be located with greater reliability since our method has no model assumptions that can be violated. These additional techniques were not investigated as a protocol relying only on DCE-MRI was desired.

15.4 Conclusions

A non-parametric variance analysis technique has been developed which appears suitable for spatially localizing parenchyma using stimulation induced concurrent with imaging. Using this technique, differences in response were noted across parotid, masseter, and pharyngeal tissues. Both intra- and inter-parotid differences were observed, and a mapping procedure was developed to quantify intra-organ differences. It is hoped that this imaging protocol (or a variation upon it) may ultimately be useful in non-invasively locating

parenchyma in head-and-neck cancer patients prior to radiotherapy so that they can be spared. It is believed this would significantly reduce toxicity risk.

Chapter 16

Conclusions

The broad aim of this thesis was to improve knowledge of late salivary gland toxicity risks for head-and-neck cancer patients treated with radiotherapy. More specifically, this work sought to: demonstrate the existence of regional effects involving late salivary dysfunction and regional radiation dose within the parotid gland, characterize the clustering nature which might lead to critical regions, and finally quantify the relative importance of whole parotid. Demonstration of similar effects in other salivary organs or using other facets was a secondary goal.

Regional effects were demonstrated. Using a careful comparative approach based on equal-volume segmentation, parametric methods that established the existence of regional effects and also provided a link between relative importance and clinical relevance, and non-parametric methods that scaled to accommodate fine segmentation, regional effects were also quantified. Limiting radiation dose to the caudal aspects of the parotid is most important for curbing dysfunction. They may also be most important for xerostomia. Some clustering was noted, though it is somewhat broad and mostly spread along the anterior-posterior direction. Submandibulars may demonstrate an inverted (in the superior-inferior direction) importance profile, but more careful investigation is needed.

16.1 Summary of Contributions

1. Systematically demonstrated the existence of a regional effect in parotid using stimulated whole-mouth saliva and sub-segment radiation doses.
2. Elucidated the clustering nature of parotid dysfunction regional effects. Demonstrated that tightly-confined clusters do not appear at the population-level[1]. Rather, regions with elevated importance are somewhat broad and concentrated in the caudal aspect.
3. Quantified parotid dysfunction regional effects throughout whole parotids, demonstrating that some regions have $>4\times$ the importance than they would if the parotid were a pure parallel organ.
4. Developed a non-invasive imaging protocol that appears able to locate salivary gland parenchyma and may be suitable for patient-specific toxicity risk assessment.
5. Demonstrated a regional effect in submandibulars (for dysfunction) and again in parotids (for xerostomia) that broadly coincides with parotid-dysfunction regional effects.
6. Devised both intra- and inter-analysis uncertainty estimation techniques based on congruence of unrelated importance methods and the emergence of smooth spatial gradients from spatially-unaware analyses. These methods are statistically underpowered, but remain viable when more sophisticated methods fail.
7. Developed a spatially unbiased segmentation procedure.
8. Extended a computational system (`DICOMautomaton`) to perform robust, iterated vector contour segmentation using a branch-and-bound technique to achieve segmentation with arbitrary volumetric constraints.

[1] It is possible that one or more confined critical regions are present within *individual* glands, however the somewhat anatomically-adaptive segmentation employed in this work suggests that such regions, if present, are patient-specific or irregularly scattered.

16.2 Avenues for Future Research

Though the findings described in this work are internally consistent, even across analysis methods and different toxicity facets, they are only in moderate agreement with the existing literature. The most promising and conclusive means of extending this research would be development of a clinical trial (or pilot study) in which the relative importance of parotid glands is accounted for during treatment planning. There are, however, a few loose ends that should first be addressed.

First, all importances derived in this work are *population*-level. It remains to be seen how applicable they are at the *individual patient*-level. A promising technique based on DCE-MRI was described in chapter 15, but it was developed wholly on volunteers. The relevance to patients receiving radiotherapy remains unknown, and the link between parenchyma and relative importance also remains unknown. A number of promising imaging techniques are emerging or have emerged in recent years (see section 3.1) – any may be suitable to elucidate the link between parenchyma and importance. Whichever method is successful, if any are, could provide a pathway to patient-specific dysfunction or xerostomia risk assessment if it can be introduced into regular clinical practice. Justification in the form of salivary organ sub-structure sparing may provide an impetus to get around the chicken-and-egg problem.

Second, the methods developed in this work are general and could in principle be applied to any OAR. As described in section 13.4, even complications like the bath-and-shower effect are thought to be handled through use of permutation importances. While some limited follow-up analysis was described in chapter 14, a more thorough analysis in other OAR is also possible.

Third, this thesis focused on the loss of salivary function. Given exciting developments linking functional recovery and stem/progenitor cells, it would be worthwhile to adapt the developed methods to try isolate regions most important for recovery. If the association between local density of stem/progenitor cells and recovery was quantified, then treatment plans could make use of dose profiles that were sculpted to promote recovery of function

in addition to loss of function. It might also suggest the most effective regions to deposit transplanted cells, as discussed in section 8.2.

Finally, simultaneous analysis of multiple facets is not currently possible using the methods developed in this thesis. Extension to accommodate multiple facets would lead to a more thorough analysis, such as derivation of inter-organ relative importance by simultaneously incorporating parotids, submandibulars, and oral cavity contributions to whole-mouth saliva would be possible with little additional work. Conversely, the simultaneous handling of multiple response variables in a single, comprehensive analysis, such as dysfunction *and* xerostomia, is an attractive end-goal with clear clinical relevance but a much clear analytical pathway. At the moment it is possible to simply combine relative importance maps from separate analysis, but the robustness of this approach is suspect.

Given that these four loose ends are all tractable, there is no practical barrier to translation of this work into a clinical pilot study. In particular, because the caudal aspect of the parotid appears to control the majority of toxicity risk, importance maps may not even need to be directly introduced into the treatment planning workflow. Rather, splitting parotid sparing practices so that some dose from the caudal-most 40-50%-volume is shifted cranially (which maintaining nodal doses on the inferior surface) may suffice to demonstrate clinically-meaningful reduction in late toxicities. Taking this 'slow introduction' approach, it may even be possible (and prudent, for comparison purposes) to maintain existing clinical mean dose thresholds.

Bibliography

[1] Haley Clark. On the regional dose susceptibility of parotid gland function loss and recovery: an effort toward amelioration of radiotherapy-induced xerostomia. Master's thesis, University of British Columbia, October 2013. URL https://open.library.ubc.ca/cIRcle/collections/24/items/1.0085551. [Accessed 25 December 2016]. (Cited on pages iv and v.)

[2] Carl Toldt and Alois Dalla Rosa. *An atlas of human anatomy for students and physicians*, volume 3-4. Rebman Company, New York, 1919. URL http://archive.org/details/atlasofhumananat34told. Alternate location: http://www.biodiversitylibrary.org/item/95995 [Accessed 9th December 2016]. (Cited on pages v, xvii, xvii, 9, and 13.)

[3] Haley D. Clark, Vitali V. Moiseenko, Thomas P. Rackley, Steven D. Thomas, Jonn S. Wu, and Stefan A. Reinsberg. Development of a method for functional aspect identification in parotid using dynamic contrast-enhanced magnetic resonance imaging and concurrent stimulation. *Acta Oncologica*, 54(9):1686–1690, 2015. doi: 10.3109/0284186X.2015.1067718. URL http://dx.doi.org/10.3109/0284186X.2015.1067718. PMID: 26252349. (Cited on pages vi, xxvii, xxvii, xxvii, xxviii, xxviii, 187, 201, 208, 209, 210, 211, and 214.)

[4] Timothy Beale and Gitta Madani. Anatomy of the salivary glands. *Seminars in Ultrasound, CT and MRI*, 27(6):436–439, 2006. ISSN 0887-2171. doi: 10.1053/j.sult.2006.09.001. URL http://dx.doi.org/10.1053/j.sult.2006.09.001. (Cited on pages xvii, 8, 8, 9, 10, 11, 11, 11, 12, 12, 12, 13, and 14.)

[5] Mastakashima Wein. *Salivary gland diseases: Surgical and medical management*. Wiley Online Library, first edition, 2007. (Cited on pages xvii and 9.)

[6] Mark Meyer, Mathieu Desbrun, Peter Schröder, and Alan H Barr. Discrete differential-geometry operators for triangulated 2-manifolds. In *Visualization and mathematics III*, pages 35–57. Springer, 2003. (Cited on pages xix, 47, and 49.)

[7] H Clark, S Thomas, V Moiseenko, R Lee, B Gill, C Duzenli, and J Wu. Automated segmentation and dose-volume analysis with DICOMautomaton. *Journal of Physics: Conference Series*, 489(1): 012009, 2014. doi: 10.1088/1742-6596/489/1/012009. URL http://dx.doi.org/10.1088/1742-6596/489/1/012009. http://stacks.iop.org/1742-6596/489/i=1/a=012009. (Cited on pages xx, xx, xxi, xxi, 55, 55, 56, 57, 57, 58, 116, 119, 126, 131, 140, 143, 173, and 205.)

[8] Simon J Sheather and Michael C Jones. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 683–690, 1991. (Cited on pages xxi, xxii, xxiii, 83, 84, and 153.)

[9] Alison C Cullen and H Christopher Frey. *Probabilistic techniques in exposure assessment: a handbook for dealing with variability and uncertainty in models and inputs*. Springer Science & Business Media, 1999. (Cited on pages xxiv, 152, and 155.)

[10] Lars Franzén, Ulrika Funegård, Thorild Ericson, and Roger Henriksson. Parotid gland function during and following radiotherapy of malignancies in the head and neck: a consecutive study of salivary flow and patient discomfort. *European Journal of Cancer*, 28(2):457–462, 1992. (Cited on pages 1, 67, 71, and 81.)

[11] M Kara Bucci, Alison Bevan, and Mack Roach. Advances in radiation therapy: conventional to 3D, to IMRT, to 4D, and beyond. *CA: a cancer journal for clinicians*, 55(2):117–134, 2005. (Cited on page 2.)

[12] Vincent Gregoire, Johannes A. Langendijk, and Sandra Nuyts. Advances in radiotherapy for head and neck cancer. *Journal of Clinical Oncology*, 33(29):3277–3284, 2015. doi: 10.1200/JCO.2015.61.2994. URL http://ascopubs.org/doi/abs/10.1200/JCO.2015.61.2994. (Cited on page 2.)

[13] R Sankaranarayanan, E Masuyer, R Swaminathan, J Ferlay, and S Whelan. Head and neck cancer: a global perspective on

epidemiology and prognosis. *Anticancer research*, 18(6B):4779–4786, 1997. (Cited on pages 2, 63, 64, 64, 64, and 109.)

[14] L Kachuri, P De, LF Ellison, R Semenciw, et al. Cancer incidence, mortality and survival trends in canada, 1970-2007. *Chronic diseases and injuries in Canada*, 33(2):69–80, 2013. (Cited on page 2.)

[15] SCH Cheng, VWC Wu, DLW Kwong, and MTC Ying. Assessment of post-radiotherapy salivary glands. *British Journal of Radiology*, 84 (1001):393–402, 2011. (Cited on pages 6, 7, 11, 11, 11, 12, 12, 13, 14, 14, 18, and 71.)

[16] Gordon B Proctor and Guy H Carpenter. Regulation of salivary gland function by autonomic nerves. *Autonomic Neuroscience*, 133(1):3–18, 2007. (Cited on pages 6, 8, 8, 11, 16, 16, 16, 76, and 77.)

[17] Henri Milne-Edwards. *Leçons sur la physiologie et l'anatomie comparée de l'homme et des animaux...*, volume 6. Librairie de Victor Masson, 1860. (Cited on pages 6 and 6.)

[18] AS Tucker. Salivary gland adaptations: Modifications of the glands for novel uses. In AS Tucker and I Miletich, editors, *Salivary Glands: Development, Adaptations, and Disease*, chapter 2, pages 21–31. Karger Publishers, London, 2010. (Cited on page 6.)

[19] JR Garrett and LC Anderson. Rat sublingual salivary glands: secretory changes on parasympathetic or sympathetic nerve stimulation and a reappraisal of the adrenergic innervation of striated ducts. *Archives of oral biology*, 36(9):675–683, 1991. (Cited on page 6.)

[20] JR Garrett. The proper role of nerves in salivary secretion: a review. *Journal of Dental Research*, 66(2):387–397, 1987. (Cited on page 6.)

[21] KL Wells. Lumen formation in salivary gland development. In AS Tucker and I Miletich, editors, *Salivary Glands: Development, Adaptations, and Disease*, chapter 5, pages 78–89. Karger Publishers, London, 2010. (Cited on page 7.)

[22] Quinn L Deveraux, Henning R Stennicke, Guy S Salvesen, and John C Reed. Endogenous inhibitors of caspases. *Journal of clinical immunology*, 19(6):388–398, 1999. (Cited on page 7.)

[23] Mitchell Vamos, Kate Welsh, Darren Finlay, Pooi San Lee, Peter D Mace, Scott J Snipas, Monica L Gonzalez, Santhi Reddy Ganji, Robert J Ardecky, Stefan J Riedl, et al. Expedient synthesis of highly potent antagonists of inhibitor of apoptosis proteins (IAPs) with unique selectivity for ML-IAP. *ACS chemical biology*, 8(4):725–732, 2013. (Cited on page 7.)

[24] I Miletich. Introduction to salivary glands: structure, function and embryonic development. In AS Tucker and Miletich I, editors, *Salivary Glands: Development, Adaptations, and Disease*, chapter 1, pages 1–20. Karger Publishers, London, 2010. (Cited on pages 7, 7, 8, 8, 8, 14, 14, 14, 14, and 15.)

[25] C Dawes. Circadian rhythms in human salivary flow rate and composition. *The Journal of physiology*, 220(3):529–545, 1972. (Cited on pages 7, 16, 71, 85, 86, and 215.)

[26] Angel I Blanco, KS Chao, Issam El Naqa, Gregg E Franklin, Konstantin Zakarian, Milos Vicic, and Joseph O Deasy. Dose–volume modeling of salivary function in patients with head-and-neck cancer receiving radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 62(4):1055–1069, 2005. (Cited on pages 7, 71, 77, 85, 89, 89, 90, 96, 99, 100, 104, 104, 111, and 149.)

[27] JA Young and E Schögel. Micropuncture investigation of sodium and potassium excretion in rat submaxillary saliva. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, 291(1):85–98, 1966. (Cited on page 8.)

[28] Mark D Sternlicht, Paul Kedeshian, Zhi-Ming Shao, Suzi Safarians, and Sanford H Barsky. The human myoepithelial cell is a natural tumor suppressor. *Clinical Cancer Research*, 3(11):1949–1958, 1997. (Cited on page 8.)

[29] Sigurd Rauch. *Die Speicheldrüsen des Menschen: Anatomie, Physiologie und klinische Pathologie*. Georg Thieme, 1959. (Cited on page 8.)

[30] G Seifert. Spezielle pathologische anatomie. *Mundhöhle, Mundspeicheldrüsen, Tonsillen und Rachen*, pages 257–259, 1966. (Cited on page 8.)

[31] Philip A Levin, James M Falko, Katharine Dixon, Elizabeth M Gallup, and William Saunderd. Benign parotid enlargement in bulimia. *Annals of internal medicine*, 93(6):827–829, 1980. (Cited on page 8.)

[32] Dorthea A Johnson. Changes in rat parotid salivary proteins associated with liquid diet-induced gland atrophy and isoproterenol-induced gland enlargement. *Archives of oral biology*, 29 (3):215–221, 1984. (Cited on pages 8 and 76.)

[33] Henry Gray. *Anatomy, descriptive and surgical*. Lea Bros., 1897. (Cited on pages 11 and 11.)

[34] Frederick S Rosen and Byron J Bailey. Anatomy and physiology of the salivary glands. *Grand rounds presentation. UTMB, Otolaryngology*, 2001. (Cited on pages 12, 14, 15, and 86.)

[35] Kirstie A Lawson. Morphogenesis and functional differentiation of the rat parotid gland in vivo and in vitro. *Journal of embryology and experimental morphology*, 24(2):411–424, 1970. (Cited on page 13.)

[36] Lawrence A Tabak. In defense of the oral cavity: structure, biosynthesis, and function of salivary mucins. *Annual review of physiology*, 57(1):547–564, 1995. (Cited on pages 14 and 15.)

[37] James L Hiatt and John J Sauk. Embryology and anatomy of the salivary glands. *Surgical Pathology of the Salivary Gland. Philadelphia: Saunders*, pages 2–9, 1991. (Cited on page 14.)

[38] David K White, H Christian Davidson, H Ric Harnsberger, Jeffrey Haller, and Aya Kamya. Accessory salivary tissue in the mylohyoid boutonnière: a clinical and radiologic pseudolesion of the oral cavity. *American journal of neuroradiology*, 22(2):406–412, 2001. (Cited on page 14.)

[39] Gary R Warnock, Jerald L Jensen, and Frank J Kratochvil. Developmental diseases. *Surgical Pathology of the Salivary Glands. Philadelphia, Saunders*, pages 10–25, 1991. (Cited on page 14.)

[40] Avraham Eisbruch, Hyungjin M Kim, Jeffrey E Terrell, Lon H Marsh, Laura A Dawson, and Jonathan A Ship. Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. *International Journal of Radiation Oncology Biology Physics*, 50(3): 695–704, 2001. (Cited on pages 14, 67, 67, 71, 71, 78, 90, 95, 95, 96, 96, 96, and 97.)

[41] L Marmiroli, G Salvi, A Caiazza, L Di Rienzo, M Massaccesi, P Murino, and G Macchia. Dose and volume impact on radiation-induced xerostomia. *Rays*, 30(2):145–148, 2004. (Cited on page 15.)

[42] AV Amerongen and ECI Veerman. Saliva–the defender of the oral cavity. *Oral diseases*, 8(1):12–22, 2002. (Cited on page 15.)

[43] Jorma Tenovuo. Antimicrobial function of human saliva-how important is it for oral health? *Acta Odontologica*, 56(5):250–256, 1998. (Cited on page 15.)

[44] Léon CPM Schenkels, Enno CI Veerman, and Arie V Nieuw Amerongen. Biochemical composition of human saliva in relation to other mucosal fluids. *Critical Reviews in Oral Biology & Medicine*, 6 (2):161–175, 1995. (Cited on page 15.)

[45] Michael Mathews, Hong Peng Jia, Janet M Guthmiller, Garrett Losh, Scott Graham, Georgia K Johnson, Brian F Tack, and Paul B McCray. Production of β-defensin antimicrobial peptides by the oral mucosa and salivary glands. *Infection and immunity*, 67(6):2740–2745, 1999. (Cited on page 15.)

[46] GB Proctor, GH Carpenter, A Segawa, JR Garrett, and L Ebersole. Constitutive secretion of immunoglobulin a and other proteins into lumina of unstimulated submandibular glands in anaesthetised rats. *Experimental physiology*, 88(1):7–12, 2003. (Cited on page 15.)

[47] DA Mackie and RM Pangborn. Mastication and its influence on human salivary flow and alpha-amylase secretion. *Physiology & behavior*, 47(3):593–595, 1990. (Cited on page 15.)

[48] P Gjörstrup. Parotid secretion of fluid and amylase in rabbits during feeding. *The Journal of physiology*, 309(1):101–116, 1980. (Cited on page 15.)

[49] Urs M Nater, Nicolas Rohleder, Jens Gaab, Simona Berger, Andreas Jud, Clemens Kirschbaum, and Ulrike Ehlert. Human salivary alpha-amylase reactivity in a psychosocial stress paradigm. *International Journal of Psychophysiology*, 55(3):333–342, 2005. (Cited on page 16.)

[50] Nils Emmelin. Control of salivary glands. *Oral Physiology, Pergamon Press, Oxford*, pages 1–16, 1972. (Cited on page 16.)

[51] Eleftheria Astreinidou, Judith M Roesink, Cornelis PJ Raaijmakers, Lambertus W Bartels, Theo D Witkamp, Jan JW Lagendijk, and Chris HJ Terhaard. 3D MR sialography as a tool to investigate radiation-induced xerostomia: feasibility study. *International Journal of Radiation Oncology Biology Physics*, 68(5):1310–1319, 2007. (Cited on page 18.)

[52] ODED Nahlieli. Classic approaches to sialoendoscopy for treatment of sialolithiasis. *Salivary Gland Diseases–Surgical and Medical Management. New York: Thieme*, pages 79–93, 2005. (Cited on pages 18, 18, 19, and 20.)

[53] Maria Lazaridou, Christos Iliopoulos, Kostas Antoniades, Ioannis Tilaveridis, Ioannis Dimitrakopoulos, and Nicolas Lazaridis. Salivary gland trauma: a review of diagnosis and treatment. *Craniomaxillofacial Trauma and Reconstruction*, 5(04):189–196, 2012. (Cited on page 18.)

[54] Reginald T Payne. Sialography: its technique and applications. *British Journal of Surgery*, 19(73):142–148, 1931. (Cited on pages 18, 18, and 19.)

[55] Haddow M Keith. Injection of the parotid gland with iodized oil. *Journal of the American Medical Association*, 90(16):1270–1271, 1928. (Cited on page 18.)

[56] G Stewart Blair. Hydrostatic sialography: An analysis of a technique. *Oral Surgery, Oral Medicine, Oral Pathology*, 36(1):116–130, 1973. (Cited on page 18.)

[57] A El-Hadary and A Ruprecht. Long-term retention of contrast medium in sialography: a case report. *Dentomaxillofacial Radiology*, 15(1):41–44, 1986. (Cited on pages 18 and 19.)

[58] RK Ngu, JE Brown, EJ Whaites, NA Drage, SY Ng, and J Makdissi. Salivary duct strictures: nature and incidence in benign salivary obstruction. *Dentomaxillofacial Radiology*, 2014. (Cited on page 18.)

[59] GJ Bansal. Digital radiography. a comparison with modern conventional imaging. *Postgraduate medical journal*, 82(969):425–428, 2006. (Cited on page 19.)

[60] Yasuro Yoshimura, Yukio Inoue, and Tetsuya Odagawa. Sonographic examination of sialolithiasis. *Journal of oral and maxillofacial surgery*, 47(9):907–912, 1989. (Cited on page 19.)

[61] Stuti Bhargava, Mukta Bhagwandas Motwani, and Vinod Madan Patni. Effect of handheld mobile phone use on parotid gland salivary flow rate and volume. *Oral surgery, oral medicine, oral pathology and oral radiology*, 114(2):200–206, 2012. (Cited on pages 19 and 85.)

[62] David M Yousem, Michael A Kraut, and Ara A Chalian. Major salivary gland imaging 1. *Radiology*, 216(1):19–29, 2000. (Cited on page 19.)

[63] Jean Provost, Clement Papadacci, Juan Esteban Arango, Marion Imbault, Mathias Fink, Jean-Luc Gennisson, Mickael Tanter, and Mathieu Pernot. 3d ultrafast ultrasound imaging in vivo. *Physics in medicine and biology*, 59(19):L1, 2014. (Cited on page 19.)

[64] Ling Tong, Hang Gao, and Jan D'hooge. Multi-transmit beam forming for fast cardiac imaging-a simulation study. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 60(8):1719–1731, 2013. (Cited on page 19.)

[65] Babak Mohammadzadeh Asl and Ali Mahloojifar. A low-complexity adaptive beamformer for ultrasound imaging using structured covariance matrix. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 59(4):660–667, 2012. (Cited on page 19.)

[66] Yipeng Lu, Hao-Yen Tang, Stephanie Fung, Bernhard E Boser, and David A Horsley. Short-range and high-resolution ultrasound imaging using an 8 mhz aluminum nitride PMUT array. In *2015 28th IEEE International Conference on Micro Electro Mechanical Systems (MEMS)*, pages 140–143. IEEE, 2015. (Cited on page 19.)

[67] Mafalda Correia, Jean Provost, Mickael Tanter, and Mathieu Pernot. 4d ultrafast ultrasound flow imaging: in vivo quantification of arterial volumetric flow rate in a single heartbeat. *Physics in Medicine and Biology*, 61(23):L48, 2016. (Cited on page 19.)

[68] Issa Loutfi, Madhusoodanan K Nair, and Ali K Ebrahim. Salivary gland scintigraphy: the use of semiquantitative analysis for uptake and clearance. *Journal of nuclear medicine technology*, 31(2):81–85, 2003. (Cited on page 19.)

[69] Simon Buus, Cai Grau, Ole Lajord Munk, Anders Rodell, Kenneth Jensen, Kim Mouridsen, and Susanne Keiding. Individual radiation response of parotid glands investigated by dynamic ^{11}C-methionine PET. *Radiotherapy and Oncology*, 78(3):262–269, 2006. (Cited on pages 19, 20, 212, 215, and 216.)

[70] Mark Kaschwich, Ulf Lützen, Yi Zhao, Angelina Tjiong, Marlies Marx, Sierk Haenisch, Oliver Wiedow, Stefanie Preuss, Juraj Culman, and Maaz Zuhayra. Biodistribution and pharmacokinetics of the 99m tc labeled human elastase inhibitor, elafin, in rats. *Drug metabolism and pharmacokinetics*, 31(2):146–155, 2016. (Cited on page 19.)

[71] Colbey Freeman, Lucia Flors, Talissa Altes, Eduard de Lange, Luke Lancaster, James Patrie, Sean Corbett, Carlos Leiva-Salinas, and Patrice Rehm. Non-contrast MRI as an alternative to Tc-99m DMSA scintigraphy in the detection of pediatric renal scarring. *Journal of Nuclear Medicine*, 57(supplement 2):1789–1789, 2016. (Cited on page 19.)

[72] Jeong-Seok Choi, Hyun-Gon Lim, Young-Mo Kim, Myung Kwan Lim, Ha Young Lee, and Jae-Yol Lim. Usefulness of magnetic resonance sialography for the evaluation of radioactive iodine-induced sialadenitis. *Annals of Surgical Oncology*, 22(3):1007–1013, 2015. (Cited on pages 20 and 20.)

[73] Nico R Mollet, Filippo Cademartiri, Carlos AG van Mieghem, Giuseppe Runza, Eugène P McFadden, Timo Baks, Patrick W Serruys, Gabriel P Krestin, and Pim J de Feyter. High-resolution spiral computed tomography coronary angiography in patients referred for diagnostic conventional coronary angiography. *Circulation*, 112(15): 2318–2323, 2005. (Cited on page 20.)

[74] Tamaki Nishi, Yasumasa Nishimura, Toru Shibata, Masaya Tamura, Naohiro Nishigaito, and Masahiko Okumura. Volume and dosimetric changes and initial clinical experience of a two-step adaptive intensity modulated radiation therapy (IMRT) scheme for head and neck cancer. *Radiotherapy and Oncology*, 106(1):85–89, 2013. (Cited on pages 20, 38, 79, 105, and 184.)

[75] A Fiorentino, R Caivano, V Metallo, C Chiumento, M Cozzolino, G Califano, S Clemente, P Pedicini, and V Fusco. Parotid gland volumetric changes during intensity-modulated radiotherapy in head

and neck cancer. *The British journal of radiology*, 2014. (Cited on pages 20 and 38.)

[76] A Wada, N Uchida, M Yokokawa, T Yoshizako, and H Kitagaki. Radiation-induced xerostomia: objective evaluation of salivary gland injury using MR sialography. *American Journal of Neuroradiology*, 30 (1):53–58, 2009. (Cited on page 20.)

[77] Oded Nahlieli and Abraham M Baruchin. Sialoendoscopy: three years' experience as a diagnostic and treatment modality. *Journal of oral and maxillofacial surgery*, 55(9):912–918, 1997. (Cited on page 20.)

[78] Oded Nahlieli, Liat Hecht Nakar, Yaron Nazarian, and Michael D Turner. Sialoendoscopy: a new approach to salivary gland obstructive pathology. *The Journal of the American Dental Association*, 137(10): 1394–1400, 2006. (Cited on page 20.)

[79] Ahmed Abdel Khalek Abdel Razek and Harish Poptani. MR spectrsocopy of head and neck cancer. *European journal of radiology*, 82(6):982–989, 2013. (Cited on page 20.)

[80] Denis Le Bihan. Apparent diffusion coefficient and beyond: what diffusion MR imaging can tell us about tissue structure. *Radiology*, 268 (2):318–322, 2013. (Cited on pages 20 and 216.)

[81] Roel GJ Kierkels, Walter H Backes, Marco HM Janssen, Jeroen Buijsen, Regina GH Beets-Tan, Philippe Lambin, Guido Lammering, Michel C Oellers, and Hugo JWL Aerts. Comparison between perfusion computed tomography and dynamic contrast-enhanced magnetic resonance imaging in rectal cancer. *International Journal of Radiation Oncology* Biology* Physics*, 77(2):400–408, 2010. (Cited on pages 20 and 213.)

[82] Amita Shukla-Dave, Nancy Y Lee, Jacobus FA Jansen, Howard T Thaler, Hilda E Stambuk, Matthew G Fury, Snehal G Patel, Andre L Moreira, Eric Sherman, Sasan Karimi, et al. Dynamic contrast-enhanced magnetic resonance imaging as a predictor of outcome in head-and-neck squamous cell carcinoma patients with nodal metastases. *International Journal of Radiation Oncology* Biology* Physics*, 82(5):1837–1844, 2012. (Cited on page 20.)

[83] Alexey Surov, Hans Jonas Meyer, Matthias Gawlitza, Anne-Kathrin Höhn, Andreas Boehm, Thomas Kahn, and Patrick Stumpp.

Correlations between DCE-MRI and histopathological parameters in head and neck squamous cell carcinoma. *Translational Oncology*, 10(1): 17–21, 2017. (Cited on page 20.)

[84] Jing Yuan, Shuzhong Chen, Ann D King, Jinyuan Zhou, Kunwar S Bhatia, Qinwei Zhang, David Ka Wei Yeung, Juan Wei, Greta Seng Peng Mok, and Yi-Xiang Wang. Amide proton transfer-weighted imaging of the head and neck at 3 t: a feasibility study on healthy human subjects and patients with head and neck cancer. *NMR in Biomedicine*, 27(10):1239–1247, 2014. (Cited on page 20.)

[85] Jing Yuan, Gladys Lo, and Ann D King. Functional magnetic resonance imaging techniques and their development for radiation therapy planning and monitoring in the head and neck cancers. *Quantitative Imaging in Medicine and Surgery*, 6(4):430, 2016. (Cited on page 20.)

[86] Matthias Gawlitza, Sandra Purz, Klaus Kubiessa, Andreas Boehm, Henryk Barthel, Regine Kluge, Thomas Kahn, Osama Sabri, and Patrick Stumpp. In vivo correlation of glucose metabolism, cell density and microcirculatory parameters in patients with head and neck cancer: initial results using simultaneous PET/MRI. *PloS one*, 10(8): e0134749, 2015. (Cited on pages 20 and 20.)

[87] Ramez Antakia, Brian H Brown, Peter E Highfield, Tim J Stephenson, Nicola J Brown, and Sabapathy P Balasubramanian. Electrical impedance spectroscopy to aid parathyroid identification and preservation in central compartment neck surgery a proof of concept in a rabbit model. *Surgical innovation*, page 1553350615607639, 2015. (Cited on page 20.)

[88] AA Danilov, VK Kramarenko, DV Nikolaev, SG Rudnev, V Yu Salamatova, AV Smirnov, and Yu V Vassilevski. Sensitivity field distributions for segmental bioelectrical impedance analysis based on real human anatomy. In *Journal of Physics: Conference Series*, volume 434, page 012001. IOP Publishing, 2013. (Cited on page 20.)

[89] K Kohli, F Liu, and K Krishnan. Validation of multi-frequency electrical impedance tomography using computed tomography. *Medical Physics*, 41(6):142–142, 2014. (Cited on page 20.)

[90] Brige Paul Chugh, Kalpagam Krishnan, Jeff Liu, and Kirpal Kohli. Evaluation of electrical impedance and computed tomography fusion

algorithms using an anthropomorphic phantom. *Medical Physics*, 41 (8):8–8, 2014. (Cited on page 20.)

[91] Yu-Xiong Su, Gui-Qing Liao, Zhuang Kang, and Yan Zou. Application of magnetic resonance virtual endoscopy as a presurgical procedure before sialoendoscopy. *The Laryngoscope*, 116(10):1899–1906, 2006. (Cited on page 20.)

[92] Yan-De Ren, Xiang-Rong Li, Jing Zhang, Li-Ling Long, Wei-Xiong Li, and Yao-Qi Han. Conventional MRI techniques combined with MR sialography on T2-3D-DRIVE in Sjögren syndrome. *International journal of clinical and experimental medicine*, 8(3):3974, 2015. (Cited on page 21.)

[93] Flavie Frémondière, Franck Lacoeuille, Avigaelle Sher, Olivier Couturier, Philippe Menei, Isabelle Richard, and Mickaël Dinomais. Isotopic scintigraphy coupled with computed tomography for the investigation of intrathecal baclofen device malfunction. *Archives of Physical Medicine and Rehabilitation*, 97(4):646–649, 2016. (Cited on page 21.)

[94] Rebecca Smith-Bindman, Diana L Miglioretti, and Eric B Larson. Rising use of diagnostic medical imaging in a large integrated health system. *Health Affairs*, 27(6):1491–1502, 2008. (Cited on page 21.)

[95] Rebecca Smith-Bindman, Diana L Miglioretti, Eric Johnson, Choonsik Lee, Heather Spencer Feigelson, Michael Flynn, Robert T Greenlee, Randell L Kruger, Mark C Hornbrook, Douglas Roblin, et al. Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010. *Jama*, 307(22):2400–2409, 2012. (Cited on page 21.)

[96] Ana CM Dovales, Luiz AR da Rosa, Ausrele Kesminiene, Mark S Pearce, and Lene HS Veiga. Patterns and trends of computed tomography usage in outpatients of the brazilian public healthcare system, 2001–2011. *Journal of Radiological Protection*, 36(3):547, 2016. (Cited on page 21.)

[97] R Farjam, N Tyagi, H Veeraraghavan, A Apte, K Zakian, M Hunt, and J Deasy. Atlas-based algorithms with local registration-goodness weighting for MRI-driven electron density mapping. *Medical Physics*, 43(6):3733–3733, 2016. (Cited on page 21.)

[98] S Ren, W Hara, Q Le, L Wang, L Xing, and R Li. A unifying multi-atlas approach to electron density mapping using multi-parametric MRI for radiation treatment planning. *Medical Physics*, 43(6):3659–3659, 2016. (Cited on page 21.)

[99] Shupeng Chen, Hong Quan, An Qin, Seonghwan Yee, and Di Yan. MR image-based synthetic CT for IMRT prostate treatment planning and CBCT image-guided localization. *Journal of Applied Clinical Medical Physics*, 17(3), 2016. (Cited on page 21.)

[100] Anthony Doemer, Indrin J Chetty, Carri Glide-Hurst, Teamour Nurushev, David Hearshen, Milan Pantelic, Melanie Traughber, Joshua Kim, Kenneth Levin, Mohamed A Elshaikh, et al. Evaluating organ delineation, dose calculation and daily localization in an open-MRI simulation workflow for prostate cancer patients. *Radiation Oncology*, 10(1):1, 2015. (Cited on page 21.)

[101] Ge Wang, Jie Zhang, Hao Gao, Victor Weir, Hengyong Yu, Wenxiang Cong, Xiaochen Xu, Haiou Shen, James Bennett, Mark Furth, et al. Towards omni-tomography – grand fusion of multiple modalities for simultaneous interior tomography. *PloS one*, 7(6):e39700, 2012. (Cited on page 21.)

[102] Rebecca Smith-Bindman, Jafi Lipson, Ralph Marcus, Kwang-Pyo Kim, Mahadevappa Mahesh, Robert Gould, Amy Berrington De González, and Diana L Miglioretti. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. *Archives of internal medicine*, 169(22): 2078–2086, 2009. (Cited on page 21.)

[103] Steve Webb. Historical experiments predating commercially available computed tomography. *The British journal of radiology*, 65(777): 835–837, 1992. (Cited on page 22.)

[104] Johann Radon. On the determination of functions from their integral values along certain manifolds. *IEEE transactions on medical imaging*, 5(4):170–176, 1986. (Cited on page 22.)

[105] Russell M Mersereau and Alan V Oppenheim. Digital reconstruction of multidimensional signals from their projections. *Proceedings of the IEEE*, 62(10):1319–1338, 1974. (Cited on page 22.)

[106] Jiang Hsieh. Computed tomography: principles, design, artifacts, and recent advances. SPIE Bellingham, WA, 2009. (Cited on page 22.)

[107] BF Wall and D Hart. Revised radiation doses for typical x-ray examinations. *The British journal of radiology*, 70(833):437–439, 1997. (Cited on page 22.)

[108] Mannudeep K Kalra, Srinivasa Prasad, Sanjay Saini, Michael A Blake, Jose Varghese, Elkan F Halpern, James T Rhea, and James H Thrall. Clinical comparison of standard-dose and 50% reducedâĂŤdose abdominal CT: Effect on image quality. *American Journal of Roentgenology*, 179(5):1101–1106, 2002. (Cited on page 22.)

[109] Denis Tack, Viviane De Maertelaer, and Pierre Alain Gevenois. Dose reduction in multidetector CT using attenuation-based online tube current modulation. *American Journal of Roentgenology*, 181(2): 331–334, 2003. (Cited on page 22.)

[110] Dunstana R Melo, Donald L Miller, Lienard Chang, Brian Moroz, Martha S Linet, and Steven L Simon. Organ doses from diagnostic medical radiographyâĂŤtrends over eight decades (1930 to 2010). *Health Physics*, 111(3):235–255, 2016. (Cited on page 22.)

[111] Siva P Raman, Pamela T Johnson, Swati Deshmukh, Mahadevappa Mahesh, Katharine L Grant, and Elliot K Fishman. CT dose reduction applications: available tools on the latest generation of CT scanners. *Journal of the American College of Radiology*, 10(1):37–41, 2013. (Cited on page 22.)

[112] Hrvoje Lusic and Mark W Grinstaff. X-ray-computed tomography contrast agents. *Chemical reviews*, 113(3):1641–1666, 2012. (Cited on page 23.)

[113] J Mitchell, RT Knight, K Kimlin, et al. Effects of iodinated contrast media on radiation therapy dosimetry for pathologies within the thorax. *Radiographer: The Official Journal of the Australian Institute of Radiography, The*, 53(2):30, 2006. (Cited on page 23.)

[114] FK-H Lee, CC-L Chan, and C-K Law. Influence of CT contrast agent on dose calculation of intensity modulated radiation therapy plan for nasopharyngeal carcinoma. *Journal of medical imaging and radiation oncology*, 53(1):114–118, 2009. (Cited on page 23.)

[115] Youngmin Choi, Jeung-Kee Kim, Hyung-Sik Lee, Won-Joo Hur, Young-Seoub Hong, Sungkwang Park, Kijung Ahn, and Heunglae Cho. Influence of intravenous contrast agent on dose calculations of intensity modulated radiation therapy plans for head and neck cancer. *Radiotherapy and oncology*, 81(2):158–162, 2006. (Cited on page 23.)

[116] Sachiko Yamada, Takashi Ueguchi, Toshiyuki Ogata, Hirokazu Mizuno, Ryota Ogihara, Masahiko Koizumi, Takeshi Shimazu, Kenya Murase, and Kazuhiko Ogawa. Radiotherapy treatment planning with contrast-enhanced computed tomography: feasibility of dual-energy virtual unenhanced imaging for improved dose calculations. *Radiation Oncology*, 9(1):1, 2014. (Cited on page 23.)

[117] Peter A Rinck. A short history of magnetic resonance imaging. *Spectroscopy Europe*, 20(1):7–10, 2008. (Cited on pages 27 and 27.)

[118] PC LAUTERBUR. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242(190): 191, 1973. (Cited on page 27.)

[119] Paul C Lauterbur. Magnetic resonance zeugmatography. *Pure and Applied Chemistry*, 40(1-2):149–157, 1974. (Cited on page 27.)

[120] Jeremy N Robinson, Andrew Coy, Robin Dykstra, Craig D Eccles, Mark W Hunter, and Paul T Callaghan. Two-dimensional NMR spectroscopy in Earth's magnetic field. *Journal of Magnetic Resonance*, 182(2):343–347, 2006. (Cited on page 28.)

[121] Carl A Michal. A low-cost spectrometer for NMR measurements in the earth's magnetic field. *Measurement Science and Technology*, 21(10): 105902, 2010. URL http://stacks.iop.org/0957-0233/21/i=10/a=105902. (Cited on page 28.)

[122] Great Britain: National Audit Office. *Managing high value capital equipment in the NHS in England: Department of Health*. Great Britain. Parliament. House of Commons. Stationery Office, 2011. ISBN 9780102969580. (Cited on page 28.)

[123] Greg J Stanisz, Ewa E Odrobina, Joseph Pun, Michael Escaravage, Simon J Graham, Michael J Bronskill, and R Mark Henkelman. T1, T2 relaxation and magnetization transfer in tissue at 3T. *Magnetic resonance in medicine*, 54(3):507–512, 2005. (Cited on page 28.)

[124] R.M. Vavrek and N.S. Grigsby. Shielded room construction for containment of fringe magnetic fields, February 1987. URL https://www.google.com/patents/US4646046. US Patent 4,646,046. (Cited on page 28.)

[125] R Jason Stafford. High field MRI: Technology, applications, safety, and limitations. *The University of Texas MD Anderson Cancer Center*, 2004. (Cited on page 28.)

[126] X Allen Li, An Tai, Douglas W Arthur, Thomas A Buchholz, Shannon Macdonald, Lawrence B Marks, Jean M Moran, Lori J Pierce, Rachel Rabinovitch, Alphonse Taghian, et al. Variability of target and normal structure delineation for breast cancer radiotherapy: an RTOG multi-institutional and multiobserver study. *International Journal of Radiation Oncology* Biology* Physics*, 73(3):944–951, 2009. (Cited on page 34.)

[127] Claudio Fiorino, Vittorio Vavassori, Giuseppe Sanguineti, Carla Bianchi, Giovanni Mauro Cattaneo, Anna Piazzolla, and Cesare Cozzarini. Rectum contouring variability in patients treated for prostate cancer: impact on rectum dose–volume histograms and normal tissue complication probability. *Radiotherapy and oncology*, 63 (3):249–255, 2002. (Cited on page 34.)

[128] Jacqueline E Livsey, James P Wylie, Ric Swindell, Vincent S Khoo, Richard A Cowan, and John P Logue. Do differences in target volume definition in prostate cancer lead to clinically relevant differences in normal tissue toxicity? *International Journal of Radiation Oncology* Biology* Physics*, 60(4):1076–1081, 2004. (Cited on page 34.)

[129] Benjamin E Nelms, Wolfgang A Tomé, Greg Robinson, and James Wheeler. Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer. *International Journal of Radiation Oncology* Biology* Physics*, 82(1):368–378, 2012. (Cited on page 34.)

[130] Charlotte L Brouwer, Roel JHM Steenbakkers, Edwin van den Heuvel, Joop C Duppen, Arash Navran, Henk P Bijl, Olga Chouvalova, Fred R Burlage, Harm Meertens, Johannes A Langendijk, et al. 3d variation in delineation of head and neck organs at risk. *Radiation Oncology*, 7 (1):1, 2012. (Cited on page 34.)

[131] Michael G Jameson, Lois C Holloway, Philip J Vial, Shalini K Vinod, and Peter E Metcalfe. A review of methods of analysis in contouring studies for radiation oncology. *Journal of medical imaging and radiation oncology*, 54(5):401–410, 2010. (Cited on page 34.)

[132] William Y Song, Bernard Chiu, Glenn S Bauman, Michael Lock, George Rodrigues, Robert Ash, Craig Lewis, Aaron Fenster, Jerry J Battista, and Jake Van Dyk. Prostate contouring uncertainty in megavoltage computed tomography images acquired with a helical tomotherapy unit during image-guided radiation therapy. *International Journal of Radiation Oncology* Biology* Physics*, 65(2): 595–607, 2006. (Cited on page 34.)

[133] Claudio Fiorino, Michele Reni, Angelo Bolognesi, Giovanni Mauro Cattaneo, and Riccardo Calandrino. Intra-and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. *Radiotherapy and oncology*, 47(3): 285–292, 1998. (Cited on page 34.)

[134] G Gong and C Liu. The contouring error of the parotids based on the CT and MRI images in radiotherapy planning for nasopharyngeal carcinoma. *Medical Physics*, 41(6):181–181, 2014. (Cited on page 34.)

[135] Primoz Petric, Johannes Dimopoulos, Christian Kirisits, Daniel Berger, Robert Hudej, and Richard Pötter. Inter-and intraobserver variation in HR-CTV contouring: intercomparison of transverse and paratransverse image orientation in 3D-MRI assisted cervix cancer brachytherapy. *Radiotherapy and Oncology*, 89(2):164–171, 2008. (Cited on page 34.)

[136] William H Hall, Michael Guiou, Nancy Y Lee, Arthur Dublin, Samir Narayan, Srinivasan Vijayakumar, James A Purdy, and Allen M Chen. Development and validation of a standardized method for contouring the brachial plexus: preliminary dosimetric analysis among patients treated with IMRT for head-and-neck cancer. *International Journal of Radiation Oncology* Biology* Physics*, 72(5):1362–1367, 2008. (Cited on page 34.)

[137] Liza J Stapleford, Joshua D Lawson, Charles Perkins, Scott Edelman, Lawrence Davis, Mark W McDonald, Anthony Waller, Eduard Schreibmann, and Tim Fox. Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer. *International*

Journal of Radiation Oncology Biology* Physics*, 77(3):959–966, 2010. (Cited on page 34.)

[138] Vladimir Pekar, Stéphane Allaire, Arish A Qazi, John J Kim, and David A Jaffray. Head and neck auto-segmentation challenge: segmentation of the parotid glands. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2010. (Cited on pages 34 and 34.)

[139] Simon K Warfield, Kelly H Zou, and William M Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE transactions on medical imaging*, 23(7):903–921, 2004. (Cited on page 34.)

[140] Gary V Walker, Musaddiq Awan, Randa Tao, Eugene J Koay, Nicholas S Boehling, Jonathan D Grant, Dean F Sittig, Gary Brandon Gunn, Adam S Garden, Jack Phan, et al. Prospective randomized double-blind study of atlas-based organ-at-risk autosegmentation-assisted radiation planning in head and neck cancer. *Radiotherapy and Oncology*, 112(3):321–325, 2014. (Cited on page 34.)

[141] Jinzhong Yang, Yongbin Zhang, Lifei Zhang, and Lei Dong. Automatic segmentation of parotids from CT scans using multiple atlases. In *Medical Image Analysis for the Clinic: A Grand Challenge*, pages 323–330. MICCAI Society, Beijing, China, 2010. (Cited on page 34.)

[142] Xiaofeng Yang, Ning Wu, Guanghui Cheng, Zhengyang Zhou, S Yu David, Jonathan J Beitler, Walter J Curran, and Tian Liu. Automated segmentation of the parotid gland based on atlas registration and machine learning: a longitudinal MRI study in head-and-neck radiation therapy. *International Journal of Radiation Oncology* Biology* Physics*, 90(5):1225–1233, 2014. (Cited on page 34.)

[143] Peter WJ Voet, Maarten LP Dirkx, David N Teguh, Mischa S Hoogeman, Peter C Levendag, and Ben JM Heijmen. Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? a dosimetric analysis. *Radiotherapy and Oncology*, 98(3):373–377, 2011. (Cited on page 34.)

[144] Jerry L Barker, Adam S Garden, K Kian Ang, Jennifer C O'Daniel, He Wang, Laurence E Court, William H Morrison, David I Rosenthal,

KS Clifford Chao, Susan L Tucker, et al. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. *International Journal of Radiation Oncology* Biology* Physics*, 59(4):960–970, 2004. (Cited on page 38.)

[145] Joel Castelli, Antoine Simon, Guillaume Louvel, Olivier Henry, Enrique Chajon, Mohamed Nassef, Pascal Haigron, Guillaume Cazoulat, Juan David Ospina, Franck Jegoux, et al. Impact of head and neck cancer adaptive radiotherapy to spare the parotid glands and decrease the risk of xerostomia. *Radiation Oncology*, 10(1):1, 2015. (Cited on page 38.)

[146] Etsushi Tomitaka, Ryuji Murakami, Keiko Teshima, Tomoko Nomura, Yuji Nakaguchi, Hideki Nakayama, Mika Kitajima, Toshinori Hirai, Yushi Araki, Masanori Shinohara, et al. Longitudinal changes over 2 years in parotid glands of patients treated with preoperative 30-gy irradiation for oral cancer. *Japanese journal of clinical oncology*, page hyq236, 2011. (Cited on page 38.)

[147] Chun-Jung Juan, Cheng-Chieh Cheng, Su-Chin Chiu, Yee-Min Jen, Yi-Jui Liu, Hui-Chu Chiu, Hung-Wen Kao, Chih-Wei Wang, Hsiao-Wen Chung, Guo-Shu Huang, et al. Temporal evolution of parotid volume and parotid apparent diffusion coefficient in nasopharyngeal carcinoma patients treated by intensity-modulated radiotherapy investigated by magnetic resonance imaging: A pilot study. *PloS one*, 10(8):e0137073, 2015. (Cited on page 38.)

[148] Yasumasa Nishimura, Kiyoshi Nakamatsu, Toru Shibata, Shuichi Kanamori, Ryuta Koike, Masahiko Okumura, and Minoru Suzuki. Importance of the initial volume of parotid glands in xerostomia for patients with head and neck cancers treated with IMRT. *Japanese Journal of Clinical Oncology*, 35(7):375–379, 2005. (Cited on pages 38, 66, 79, 89, and 105.)

[149] Sue S Taji, Neil Savage, Trevor Holcombe, Farid Khan, and W Kim Seow. Congenital aplasia of the major salivary glands: literature review and case report. *Pediatric dentistry*, 33(2):113–118, 2011. (Cited on page 38.)

[150] Soon-Young Kwon, Eun-Jae Jung, Seung-Hyun Kim, and Tack-Koon

Kim. A case of major salivary gland agenesis. *Acta oto-laryngologica*, 126(2):219–222, 2006. (Cited on page 38.)

[151] Woo Kyung Moon, Moon Hee Han, In-One Kim, Myung Whun Sung, Kee Hyun Chang, Sung Wook Choo, and Man Chung Han. Congenital fistula from ectopic accessory parotid gland: diagnosis with CT sialography and CT fistulography. *American journal of neuroradiology*, 16(4):997–999, 1995. (Cited on page 38.)

[152] PH Doyle and DA Moran. A short proof that compact 2-manifolds can be triangulated. *Inventiones mathematicae*, 5(2):160–162, 1968. (Cited on page 41.)

[153] William P Thurston. Three-dimensional manifolds, kleinian groups and hyperbolic geometry. In *Proc. Sympos. Pure Math*, volume 39, pages 87–111, 1983. (Cited on page 43.)

[154] Huai-Dong Cao, Xi-Ping Zhu, et al. A complete proof of the Poincaré and geometrization conjectures-application of the Hamilton-Perelman theory of the Ricci flow. *Asian Journal of Mathematics*, 10(2):165–492, 2006. (Cited on page 43.)

[155] Giuseppe Peano. Sur une courbe, qui remplit toute une aire plane. *Mathematische Annalen*, 36(1):157–160, 1890. (In which Peano introduced space-filling curves.). (Cited on page 43.)

[156] David Hilbert. Ueber die stetige abbildung einer line auf ein flächenstück. *Mathematische Annalen*, 38(3):459–460, 1891. (In which Hilbert introduced the Hilbert curve.). (Cited on pages 43 and 44.)

[157] James J Kuffner and Steven M LaValle. Space-filling trees: A new perspective on incremental search for motion planning. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2199–2206. IEEE, 2011. (Cited on page 43.)

[158] Steven M LaValle. Rapidly-exploring random trees: A new tool for path planning. Technical report, 1998. (Cited on page 43.)

[159] Przemyslaw Prusinkiewicz and Aristid Lindenmayer. *The algorithmic beauty of plants*. Springer Science & Business Media, 2012. (Cited on page 43.)

[160] Jonathan A Quinn, Frank C Langbein, and Ralph R Martin. Low-discrepancy point sampling of meshes for rendering. In *SPBG*, pages 19–28, 2007. (Cited on pages 43 and 45.)

[161] Jonathan A. Quinn, Frank C. Langbein, Yu kun Lai, and Ralph R. Martin. Fast low-discrepancy sampling of parametric surfaces and meshes. In *Proceedings of the Conference on the Mathematics of Surfaces XIV*. IMA, 2013. doi: 10.13140/RG.2.1.2012.0408. URL http://dx.doi.org/10.13140/RG.2.1.2012.0408. (Cited on page 43.)

[162] Donghong Ding, Zengxi Pan, Dominic Cuiuri, Huijun Li, and Stephen van Duin. Advanced design for additive manufacturing: 3d slicing and 2d path planning. In Igor V Shishkovsky, editor, *New Trends in 3D Printing*, pages 1–23, 2016. doi: 10.5772/63042. URL http://dx.doi.org/10.5772/63042. (Cited on page 43.)

[163] Tobias Wasser, Anshu Dhar Jayal, and Christoph Pistor. Implementation and evaluation of novel buildstyles in fused deposition modeling (FDM). *Strain*, 5(6), 1999. (Cited on pages 43, 43, and 45.)

[164] Bongki Moon, Hosagrahar V Jagadish, Christos Faloutsos, and Joel H. Saltz. Analysis of the clustering properties of the hilbert space-filling curve. *IEEE Transactions on knowledge and data engineering*, 13(1): 124–141, 2001. (Cited on page 44.)

[165] Yaroslav D Sergeyev, Roman G Strongin, and Daniela Lera. *Introduction to global optimization exploiting space-filling curves*. Springer Science & Business Media, 2013. (Cited on page 44.)

[166] Jordan J Cox, Yasuko Takezaki, Helaman RP Ferguson, Kent E Kohkonen, and Eric L Mulkay. Space-filling curves in tool-path applications. *Computer-Aided Design*, 26(3):215–224, 1994. (Cited on page 44.)

[167] F Chauny, Alain Haurie, E Wagneur, and R Loulou. Sequencing punch operations in a flexible manufacturing cell a three-dimensional space-filling curve approach. *INFOR: Information Systems and Operational Research*, 25(1):26–45, 1987. (Cited on page 44.)

[168] E Netto. Beitrag zur mannigfaltigkeitslehre. *Crelle J*, 86:263–268, 1879. (Cited on page 45.)

[169] Guohua Jin and John Mellor-Crummey. SFCGen: a framework for efficient generation of multi-dimensional space-filling curves by recursion. *ACM Transactions on Mathematical Software (TOMS)*, 31 (1):120–148, 2005. (Cited on page 45.)

[170] JA Scott. Some examples of the use of areal coordinates in triangle geometry. *The Mathematical Gazette*, 83(498):472–477, 1999. (Cited on pages 45 and 45.)

[171] Mark Meyer, Alan Barr, Haeyoung Lee, and Mathieu Desbrun. Generalized barycentric coordinates on irregular polygons. *Journal of graphics tools*, 7(1):13–22, 2002. (Cited on page 45.)

[172] Joe Warren. Barycentric coordinates for convex polytopes. *Advances in Computational Mathematics*, 6(1):97–108, 1996. (Cited on page 45.)

[173] Joe Warren, Scott Schaefer, Anil N Hirani, and Mathieu Desbrun. Barycentric coordinates for convex sets. *Advances in computational mathematics*, 27(3):319–338, 2007. (Cited on page 45.)

[174] Eugene L Wachspress. *A rational finite element basis*. Elsevier, 1975. (Cited on page 45.)

[175] N Sukumar and EA Malsch. Recent advances in the construction of polygonal finite element interpolants. *Archives of Computational Methods in Engineering*, 13(1):129–163, 2006. (Cited on page 45.)

[176] Tao Ju, Scott Schaefer, and Joe Warren. Mean value coordinates for closed triangular meshes. In *ACM Transactions on Graphics*, volume 24, pages 561–566. ACM, 2005. (Cited on page 45.)

[177] Michael S Floater, Géza Kós, and Martin Reimers. Mean value coordinates in 3d. *Computer Aided Geometric Design*, 22(7):623–631, 2005. (Cited on page 45.)

[178] Frank Morgan. Manifolds with density and perelman's proof of the poincaré conjecture. *American Mathematical Monthly*, 116(2):134–142, 2009. (Cited on page 47.)

[179] Grigori Perelman. *Saddle surfaces in Euclidean spaces*. PhD thesis, Leningrad State University, 1990. (Cited on page 47.)

[180] Lior Shapira, Ariel Shamir, and Daniel Cohen-Or. Consistent mesh partitioning and skeletonisation using the shape diameter function. *The Visual Computer*, 24(4):249–259, 2008. (Cited on page 48.)

[181] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996. (Cited on page 50.)

[182] Jonathan Richard Shewchuk. Delaunay refinement algorithms for triangular mesh generation. *Computational geometry*, 22(1):21–74, 2002. (Cited on page 50.)

[183] CT Loop. Smooth subdivision surfaces based on triangles. Master's thesis, University of Utah, August 1987. (Cited on page 50.)

[184] W Dean Bidgood and Steven C Horii. Introduction to the ACR-NEMA DICOM standard. *Radiographics*, 12(2):345–355, 1992. (Cited on page 52.)

[185] DICOM Standards Committee Working Group 6. Dicom supplement 132: Surface segmentation storage sop class. Technical report, 2008. URL ftp://medical.nema.org/medical/dicom/final/sup132_ft.pdf. (Cited on page 52.)

[186] John Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, 2007. (Cited on page 58.)

[187] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Non-parametric diffeomorphic image registration with the demons algorithm. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 319–326. Springer, 2007. (Cited on page 58.)

[188] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Symmetric log-domain diffeomorphic registration: A demons-based approach. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 754–761. Springer, 2008. (Cited on page 58.)

[189] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72, 2009. (Cited on page 58.)

[190] Florence Dru and Tom Vercauteren. An ITK Implementation of the Symmetric Log-Domain Diffeomorphic Demons Algorithm, May 2009. URL https://hal.inria.fr/hal-00813744. (Cited on page 58.)

[191] John A Pinkston and Philip Cole. Incidence rates of salivary gland tumors: results from a population-based study. *Otolaryngology-Head and Neck Surgery*, 120(6):834–840, 1999. (Cited on page 61.)

[192] R Hornigold, PR Morgan, A Pearce, and MJ Gleeson. Congenital sialolipoma of the parotid gland first reported case and review of the literature. *International journal of pediatric otorhinolaryngology*, 69(3): 429–434, 2005. (Cited on page 61.)

[193] John G Armstrong, Louis B Harrison, Ronald H Spiro, Daniel E Fass, Elliot W Strong, and Zvi Y Fuks. Malignant tumors of major salivary gland origin: a matched-pair analysis of the role of combined surgery and postoperative radiotherapy. *Archives of OtolaryngologyâĂŤHead & Neck Surgery*, 116(3):290, 1990. (Cited on page 61.)

[194] Charles R Leemans, Rammohan Tiwari, Jos JP Nauta, Isaac Van Der Waal, and Gordon B Snow. Recurrence at the primary site in head and neck cancer and the significance of neck lymph node metastases as a prognostic factor. *Cancer*, 73(1):187–190, 1994. (Cited on page 62.)

[195] Lynn AG Ries, D Harkins, M Krapcho, Angela Mariotto, Barry A Miller, Eric J Feuer, Limin X Clegg, MP Eisner, Marie-Josèphe Horner, Nadia Howlader, et al. SEER cancer statistics review, 1975-2003. 2006. (Cited on pages 62, 63, and 64.)

[196] Wael A Sakr, Richard J Zarbo, John R Jacobs, and John D Crissman. Distribution of basement membrane in squamous cell carcinoma of the head and neck. *Human pathology*, 18(10):1043–1050, 1987. (Cited on page 62.)

[197] BA Gusterson, MJ Warburton, D Mitchell, N Kraft, and WW Hancock. Invading squamous cell carcinoma can retain a basal lamina. an immunohistochemical study using a monoclonal antibody to type IV collagen. *Laboratory investigation; a journal of technical methods and pathology*, 51(1):82–87, 1984. (Cited on page 62.)

[198] Isaiah J Fidler and Ian R Hart. Biological diversity in metastatic neoplasms: origins and implications. *Science*, 217(4564):998–1003, 1982. (Cited on pages 62 and 63.)

[199] Don X Nguyen and Joan Massagué. Genetic determinants of cancer metastasis. *Nature Reviews Genetics*, 8(5):341–352, 2007. (Cited on page 63.)

[200] Athanassios Argiris, Bruce E Brockstein, Daniel J Haraf, Kerstin M Stenson, Bharat B Mittal, Merrill S Kies, Fred R Rosen, Borko Jovanovic, and Everett E Vokes. Competing causes of death and second primary tumors in patients with locoregionally advanced head and neck cancer treated with chemoradiotherapy. *Clinical Cancer Research*, 10(6):1956–1962, 2004. (Cited on page 63.)

[201] Fadlo R Khuri, J Jack Lee, Scott M Lippman, Edward S Kim, Jay S Cooper, Steven E Benner, Rodger Winn, Thomas F Pajak, Brendell Williams, George Shenouda, et al. Randomized phase III trial of low-dose isotretinoin for prevention of second primary tumors in stage I and II head and neck cancer patients. *Journal of the National Cancer Institute*, 98(7):441–450, 2006. (Cited on pages 63 and 107.)

[202] James A Bonner, Paul M Harari, Jordi Giralt, Nozar Azarnia, Dong M Shin, Roger B Cohen, Christopher U Jones, Ranjan Sur, David Raben, Jacek Jassem, et al. Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. *New England Journal of Medicine*, 354(6):567–578, 2006. (Cited on pages 63, 109, 109, and 110.)

[203] Maria Teresa Canto and Susan S Devesa. Oral cavity and pharynx cancer incidence rates in the united states, 1975–1998. *Oral oncology*, 38(6):610–617, 2002. (Cited on page 63.)

[204] Athanassios Argiris, Michalis V Karamouzis, David Raben, and Robert L Ferris. Head and neck cancer. *The Lancet*, 371(9625): 1695–1709, 2008. (Cited on page 63.)

[205] Mur At Beyz Adeoglu, Gokh An Ozyigit, and Ugur Selek. *Radiation Therapy for Head and Neck Cancers*. Springer, 2015. (Cited on page 63.)

[206] Manabu Muto, Keiko Minashi, Tomonori Yano, Yutaka Saito, Ichiro Oda, Satoru Nonaka, Tai Omori, Hitoshi Sugiura, Kenichi Goda, Mitsuru Kaise, et al. Early detection of superficial squamous cell carcinoma in the head and neck region and esophagus by narrow band imaging: a multicenter randomized controlled trial. *Journal of Clinical Oncology*, 28(9):1566–1572, 2010. (Cited on pages 63 and 107.)

[207] Paul Brocklehurst, Omar Kujan, Lucy A O'Malley, Graham Ogden, Simon Shepherd, and Anne-Marie Glenny. Screening programmes for the early detection and prevention of oral cancer. *The Cochrane Library*, 2013. (Cited on page 63.)

[208] Vinay Kumar, Abul K Abbas, Nelson Fausto, and Jon C Aster. *Robbins and Cotran pathologic basis of disease*. Elsevier Health Sciences, 2014. (Cited on page 63.)

[209] K Thomas Robbins, Jesus E Medina, Gregory T Wolfe, Paul A Levine, Roy B Sessions, and Charles W Pruet. Standardizing neck dissection terminology: official report of the academy's committee for head and neck surgery and oncology. *Archives of Otolaryngology–Head & Neck Surgery*, 117(6):601–605, 1991. (Cited on pages 63 and 63.)

[210] James C Alex, Clarence T Sasaki, David N Krag, Barry Wenig, and Paula B Pyle. Sentinel lymph node radiolocalization in head and neck squamous cell carcinoma. *The Laryngoscope*, 110(2):198–198, 2000. (Cited on pages 64 and 64.)

[211] Ahmedin Jemal, Rebecca Siegel, Elizabeth Ward, Taylor Murray, Jiaquan Xu, and Michael J Thun. Cancer statistics, 2007. *CA: a cancer journal for clinicians*, 57(1):43–66, 2007. (Cited on page 64.)

[212] Ahmedin Jemal, Rebecca Siegel, Elizabeth Ward, Yongping Hao, Jiaquan Xu, and Michael J Thun. Cancer statistics, 2009. *CA: a cancer journal for clinicians*, 59(4):225–249, 2009. (Cited on page 64.)

[213] D Max Parkin, Freddie Bray, J Ferlay, and Paola Pisani. Global cancer statistics, 2002. *CA: a cancer journal for clinicians*, 55(2):74–108, 2005. (Cited on page 64.)

[214] Jielin Feng and Robert P Coppes. Can we rescue salivary gland function after irradiation. *Sci World J*, 8:959–962, 2008. (Cited on pages 64, 103, and 146.)

[215] Freddi Lewin, Staffan E Norell, Hemming Johansson, Per Gustavsson, Johan Wennerberg, Anders Biörklund, and Lars Erik Rutqvist. Smoking tobacco, oral snuff, and alcohol in the etiology of squamous cell carcinoma of the head and neck. *Cancer*, 82(7):1367–1375, 1998. (Cited on page 64.)

[216] Athanassios Argiris and Cathy Eng. Epidemiology, staging, and screening of head and neck cancer. In *Head and neck cancer*, pages 15–60. Springer, 2004. (Cited on page 64.)

[217] Gypsyamber D'Souza, Aimee R Kreimer, Raphael Viscidi, Michael Pawlita, Carole Fakhry, Wayne M Koch, William H Westra, and

Maura L Gillison. Case–control study of human papillomavirus and oropharyngeal cancer. *New England Journal of Medicine*, 356(19): 1944–1956, 2007. (Cited on page 64.)

[218] Mia Hashibe, Paolo Boffetta, David Zaridze, Oxana Shangina, Neonila Szeszenia-Dabrowska, Dana Mates, Vladimir Janout, Eleonóra Fabiánová, Vladimir Bencko, Norman Moullan, et al. Evidence for an important role of alcohol-and aldehyde-metabolizing genes in cancers of the upper aerodigestive tract. *Cancer Epidemiology Biomarkers & Prevention*, 15(4):696–703, 2006. (Cited on page 64.)

[219] Erich M Sturgis and Qingyi Wei. Genetic susceptibility–molecular epidemiology of head and neck cancer. *Current opinion in oncology*, 14 (3):310–317, 2002. (Cited on pages 64 and 64.)

[220] Maria Pavia, Claudia Pileggi, Carmelo GA Nobile, and Italo F Angelillo. Association between fruit and vegetable consumption and oral cancer: a meta-analysis of observational studies. *The American journal of clinical nutrition*, 83(5):1126–1134, 2006. (Cited on page 64.)

[221] Nicole K Proia, Geraldine M Paszkiewicz, Maureen A Sullivan Nasca, Gail E Franke, and John L Pauly. Smoking and smokeless tobacco-associated human buccal cell mutations and their association with oral cancer – a review. *Cancer Epidemiology Biomarkers & Prevention*, 15(6):1061–1077, 2006. (Cited on page 64.)

[222] S Warnakulasuriya. Areca nut use following migration and its consequences. *Addiction biology*, 7(1):127–132, 2002. (Cited on page 64.)

[223] MD Mignogna, S Fedele, and L Lo Russo. The world cancer report and the burden of oral cancer. *European Journal of Cancer Prevention*, 13(2):139–142, 2004. (Cited on page 64.)

[224] Cathy L Bartels. Xerostomia information for dentists. *Oral Cancer Foundation Cite*, 7, 2005. (Cited on pages 66 and 77.)

[225] Siri Jensen, Anne Pedersen, Jesper Reibel, and Birgitte Nauntofte. Xerostomia and hypofunction of the salivary glands in cancer therapy. *Supportive care in cancer*, 11(4):207–225, 2003. (Cited on pages 66, 70, and 109.)

[226] Alexander Lin, Hyungjin M Kim, Jeffrey E Terrell, Laura A Dawson, Jonathan A Ship, and Avraham Eisbruch. Quality of life after parotid-sparing IMRT for head-and-neck cancer: a prospective longitudinal study. *International Journal of Radiation Oncology Biology Physics*, 57(1):61–70, 2003. (Cited on pages 66, 69, 71, 71, and 101.)

[227] WJ Loesche, J Bromberg, MS Terpenning, WA Bretz, BL Dominguez, NS Grossman, and SE Langmore. Xerostomia, xerogenic medications and food avoidances in selected geriatric groups. *Journal of the American Geriatrics Society*, 43(4):401–407, 1995. (Cited on pages 66, 67, and 88.)

[228] TA Makkonen and E Nordman. Estimation of long-term salivary gland damage induced by radiotherapy. *Acta Oncologica*, 26(4):307–312, 1987. (Cited on page 67.)

[229] KS Chao, Joseph O Deasy, Jerry Markman, Joyce Haynie, Carlos A Perez, James A Purdy, and Daniel A Low. A prospective study of salivary function sparing in patients with head-and-neck cancers receiving intensity-modulated or three-dimensional radiation therapy: initial results. *International Journal of Radiation Oncology Biology Physics*, 49(4):907–916, 2001. (Cited on pages 67, 71, 78, 89, 96, 96, 97, and 104.)

[230] L Clifton Stephens, K Kian Ang, Timothy E Schultheiss, Glen K King, William A Brock, and Lester J Peters. Target cell and mode of radiation injury in rhesus salivary glands. *Radiotherapy and Oncology*, 7(2):165–174, 1986. (Cited on pages 67 and 78.)

[231] L Clifton Stephens, Timothy E Schultheiss, Roger E Price, K Kian Ang, and Lester J Peters. Radiation apoptosis of serous acinar cells of salivary and lacrimal glands. *Cancer*, 67(6):1539–1543, 1991. (Cited on pages 67 and 78.)

[232] Hirohiko Tsuni. Quantitative dose-response analysis of salivary function following radiotherapy using sequential RI-sialography. *International Journal of Radiation Oncology Biology Physics*, 11(9): 1603–1612, 1985. (Cited on pages 67, 78, and 170.)

[233] F Buettner, AB Miah, SL Gulliford, E Hall, KJ Harrington, S Webb, M Partridge, and CM Nutting. Novel approaches to improve the

therapeutic index of head and neck radiotherapy: an analysis of data from the PARSPORT randomised phase III trial. *Radiotherapy and Oncology*, 103(1):82, 2012. (Cited on pages 67, 88, 89, 89, 102, 105, 146, 168, 171, 172, 172, 185, and 202.)

[234] Hadi Seikaly, Naresh Jha, Timothy McGaw, Linda Coulter, Richard Liu, and Derald Oldring. Submandibular gland transfer: a new method of preventing radiation-induced xerostomia. *The Laryngoscope*, 111(2):347–352, 2001. (Cited on pages 67 and 89.)

[235] Khalid Al-Qahtani, Michael P Hier, Khalil Sultanum, and Martin J Black. The role of submandibular salivary gland transfer in preventing xerostomia in the chemoradiotherapy patient. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, 101(6): 753–756, 2006. (Cited on pages 67 and 89.)

[236] Joseph O Deasy, Vitali Moiseenko, Lawrence Marks, KS Chao, Jiho Nam, and Avraham Eisbruch. Radiotherapy dose–volume effects on salivary gland function. *International Journal of Radiation Oncology Biology Physics*, 76(3):S58–S63, 2010. (Cited on pages 67, 89, 100, 101, 111, 126, 145, 170, 171, 201, and 202.)

[237] Leo M Sreebny and Arjan Vissink. *Dry mouth, the malevolent symptom: a clinical guide*. Wiley. com, 2010. (Cited on pages 68 and 79.)

[238] A.G. Bartley. Suppression of the saliva. *Medical Times and Gazette*, 54:603, 1868. (Published as a letter to the editor). (Cited on page 68.)

[239] Jonathan Hutchinson and T Fitz-Patrick. Xerostomia – a term suggested for the affection known as dry mouth. *The Lancet*, 133 (3416):352–353, 1889. (Cited on page 68.)

[240] Hayes E Martin. Treatment of pharyngeal cancer. *Archives of Otolaryngology – Head and Neck Surgery*, 27(6):661, 1938. (Cited on pages 68, 69, and 69.)

[241] Shrinivas Rathod, Tejpal Gupta, Sarbani Ghosh-Laskar, Vedang Murthy, Ashwini Budrukkar, and JaiPrakash Agarwal. Quality-of-life (QoL) outcomes in patients with head and neck squamous cell carcinoma (HNSCC) treated with intensity-modulated radiation therapy (IMRT) compared to three-dimensional conformal

radiotherapy (3D-CRT): Evidence from a prospective randomized study. *Oral Oncology*, 2013. (Cited on pages 69 and 101.)

[242] A Vissink, FR Burlage, FKL Spijkervet, J Jansma, and RP Coppes. Prevention and treatment of the consequences of head and neck radiotherapy. *Critical Reviews in Oral Biology & Medicine*, 14(3): 213–225, 2003. (Cited on pages 69, 70, and 70.)

[243] Nishtha Gupta, Manoj Pal, Sheh Rawat, Mandeep S Grewal, Himani Garg, Deepika Chauhan, Parveen Ahlawat, Sarthak Tandon, Ruparna Khurana, Anjali K Pahuja, et al. Radiation-induced dental caries, prevention and treatment-a systematic review. *National Journal of Maxillofacial Surgery*, 6(2):160, 2015. (Cited on pages 69, 69, and 70.)

[244] Ingo N Springer, Peter Niehoff, Patrick H Warnke, Güler Böcek, György Kovács, Manfred Suhr, Jörg Wiltfang, and Yahya Açil. Radiation cariesâĂŤradiogenic destruction of dental collagen. *Oral oncology*, 41(7):723–728, 2005. (Cited on page 69.)

[245] Andrej M Kielbassa, Wolfgang Hinkelbein, Elmar Hellwig, and Hendrik Meyer-Lückel. Radiation-related damage to dentition. *The lancet oncology*, 7(4):326–335, 2006. (Cited on page 69.)

[246] Christopher G Murray, Jay Herson, Thomas E Daly, and Stuart Zimmerman. Radiation necrosis of the mandible: a 10 year study. part II. dental factors; onset, duration and management of necrosis. *International Journal of Radiation Oncology* Biology* Physics*, 6(5): 549–553, 1980. (Cited on page 69.)

[247] JA Del Regato. Dental lesions observed after roentgen therapy in cancer of the buccal cavity, pharynx and larynx. *Am J Roentgenol*, 42: 404–410, 1939. (Cited on page 70.)

[248] Oral Assessment before Treatment. Oral and dental management related to radiation therapy for head and neck cancer. *J Can Dent Assoc*, 69(9):585–90, 2003. (Cited on page 70.)

[249] Paulo Rogério Ferreti Bonan, Márcio Ajudarte Lopes, Fábio Ramoa Pires, and Oslei Paes de Almeida. Dental management of low socioeconomic level patients before radiotherapy of the head and neck with special emphasis on the prevention of osteoradionecrosis. *Brazilian dental journal*, 17(4):336–342, 2006. (Cited on page 70.)

[250] Shiyana Eliyas, Ahmed Al-Khayatt, Richard WJ Porter, and Peter Briggs. Dental extractions prior to radiotherapy to the jaws for reducing post-radiotherapy dental complications. *The Cochrane Library*, 2013. (Cited on page 70.)

[251] Joel B Epstein, Giuseppe Rea, Frances LW Wong, John Spinelli, and Peter Stevenson-Moore. Osteonecrosis: study of the relationship of dental extractions in patients receiving radiotherapy. *Head & neck surgery*, 10(1):48–54, 1987. (Cited on page 70.)

[252] Frankie Sulaiman, Joseph M Huryn, and Ian M Zlotolow. Dental extractions in the irradiated head and neck patient: a retrospective analysis of memorial sloan-kettering cancer center protocols, criteria, and end results. *Journal of oral and maxillofacial surgery*, 61(10): 1123–1131, 2003. (Cited on page 70.)

[253] WG Maxymiw, RE Wood, and F-F Liu. Postradiation dental extractions without hyperbaric oxygen. *Oral surgery, oral medicine, oral pathology*, 72(3):270–274, 1991. (Cited on page 70.)

[254] Morten Schiødt and Niels Hermund. Management of oral disease prior to radiation therapy. *Supportive Care in Cancer*, 10(1):40–43, 2002. (Cited on page 70.)

[255] Jens J Thorn, Hanne Sand Hansen, Lena Specht, and Lars Bastholt. Osteoradionecrosis of the jaws: clinical characteristics and relation to the field of irradiation. *Journal of Oral and Maxillofacial Surgery*, 58 (10):1088–1093, 2000. (Cited on page 70.)

[256] JB Epstein, S Emerton, ND Le, and P Stevenson-Moore. A double-blind crossover trial of oral balance gel and biotene® toothpaste versus placebo in patients with xerostomia following radiation therapy. *Oral oncology*, 35(2):132–137, 1999. (Cited on page 70.)

[257] NR Kressin, U Boehmer, ME Nunn, and A Spiro. Increased preventive practices lead to greater tooth retention. *Journal of dental research*, 82 (3):223–227, 2003. (Cited on page 70.)

[258] P Del Vigna de Almeida, AM Gregio, MA Machado, AA De Lima, and L Reis Azevedo. Saliva composition and functions: a comprehensive review. *J Contemp Dent Pract*, 9(3):72–80, 2008. (Cited on pages 70 and 77.)

[259] JC Atkinson and PC Fox. Salivary gland dysfunction. *Clinics in geriatric medicine*, 8(3):499–511, 1992. (Cited on page 70.)

[260] Neil P Walsh, Joanna C Montague, Nichola Callow, and Ann V Rowlands. Saliva flow rate, total protein concentration and osmolality as potential markers of whole body hydration status during progressive acute dehydration in humans. *Archives of Oral Biology*, 49(2):149–154, 2004. (Cited on page 71.)

[261] Neil P Walsh, Stewart J Laing, Samuel J Oliver, Joanna C Montague, Robert Walters, and James LJ Bilzon. Saliva parameters as potential indices of hydration status during acute dehydration. *Medicine and science in sports and exercise*, 36(9):1535–1542, 2004. (Cited on page 71.)

[262] Philip C Fox. Acquired salivary dysfunction: drugs and radiation. *Annals of the New York Academy of Sciences*, 842(1):132–137, 1998. (Cited on page 71.)

[263] PC Fox, KA Busch, and BJ Baum. Subjective reports of xerostomia and objective measures of salivary gland performance. *The Journal of the American Dental Association*, 115(4):581–584, 1987. (Cited on pages 71, 71, and 71.)

[264] Han Liem, Renato A Valdes Olmos, Alfons JM Balm, Ronald B Keus, Harm van Tinteren, Robert P Takes, Sara H Muller, Allison M Bruce, Cornelis A Hoefnagel, and Frans J Hilgers. Evidence for early and persistent impairment of salivary gland excretion after irradiation of head and neck tumours. *European Journal of Nuclear Medicine*, 23 (11):1485–1490, 1996. (Cited on page 71.)

[265] Kenneth Jensen, Karin Lambertsen, Peter Torkov, Martin Dahl, Anders Bonde Jensen, and Cai Grau. Patient assessed symptoms are poor predictors of objective findings. results from a cross sectional study in patients treated with radiotherapy for pharyngeal cancer. *Acta Oncologica*, 46(8):1159–1168, 2007. (Cited on page 71.)

[266] Kenneth Jensen, Anders Bonde Jensen, and Cai Grau. The relationship between observer-based toxicity scoring and patient assessed symptom severity after treatment for head and neck cancer. a correlative cross sectional study of the DAHANCA toxicity scoring system and the EORTC quality of life questionnaires. *Radiotherapy and Oncology*, 78(3):298–305, 2006. (Cited on pages 71, 75, and 75.)

[267] Chad M Amosson, Bin S Teh, T John Van, Nathan Uy, Eugene Huang, Wei-Yuan Mai, Anna Frolov, Shiao Y Woo, J Kam Chiu, L Steven Carpenter, et al. Dosimetric predictors of xerostomia for head-and-neck cancer patients treated with the smart (simultaneous modulated accelerated radiation therapy) boost technique. *International Journal of Radiation Oncology Biology Physics*, 56(1): 136–144, 2003. (Cited on pages 71 and 111.)

[268] Edmond HN Pow, Dora LW Kwong, Anne S McMillan, May Wong, Jonathan ST Sham, Lucullus HT Leung, and W Keung Leung. Xerostomia and quality of life after intensity-modulated radiotherapy vs. conventional radiotherapy for early-stage nasopharyngeal carcinoma: initial report on a randomized controlled clinical trial. *International Journal of Radiation Oncology* Biology* Physics*, 66(4): 981–991, 2006. (Cited on page 71.)

[269] Peter F Kohler and Margaret E Winter. A quantitative test for xerostomia. the saxon test, an oral equivalent of the schirmer test. *Arthritis & Rheumatism*, 28(10):1128–1132, 1985. (Cited on page 71.)

[270] James D Evans. *Straightforward statistics for the behavioral sciences*. Brooks/Cole, 1996. (Cited on pages 72 and 190.)

[271] Richard Taylor. Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography*, 6(1):35–39, 1990. (Cited on page 72.)

[272] Gabriel Gabriel, Michael Barton, and Geoff P Delaney. The effect of travel distance on radiotherapy utilization in NSW and ACT. *Radiotherapy and Oncology*, 117(2):386–389, 2015. (Cited on pages 74, 107, and 108.)

[273] Vitali Moiseenko, Jonn Wu, Allan Hovan, Ziad Saleh, Aditya Apte, Joseph O Deasy, Stephen Harrow, Carman Rabuka, Adam Muggli, and Anna Thompson. Treatment planning constraints to avoid xerostomia in head-and-neck radiotherapy: an independent test of QUANTEC criteria using a prospectively collected dataset. *International Journal of Radiation Oncology Biology Physics*, 82(3):1108–1114, 2012. (Cited on pages 75, 111, and 111.)

[274] Andy Trotti, Roger Byhardt, Joanne Stetz, Clement Gwede, Benjamin Corn, Karen Fu, Leonard Gunderson, Beryl McCormick, Mitchell

Morris, Tyvin Rich, et al. Common toxicity criteria: version 2.0. an improved reference for grading the acute effects of cancer treatment: impact on radiotherapy. *International Journal of Radiation Oncology* Biology* Physics*, 47(1):13–47, 2000. (Cited on page 75.)

[275] AD Colevas and A Setser. The nci Common Terminology Criteria for Adverse Events (CTCAE) version 3.0 is the new standard for oncology clinical trials. In *ASCO Annual Meeting Proceedings*, volume 22, page 6098, 2004. (Cited on page 75.)

[276] US Department of Health, Human Services, et al. Common terminology criteria for adverse events (CTCAE) version 4.0. *National Cancer Institute*, (09-5410), 2009. (Cited on page 75.)

[277] Andy Trotti, A. Dimitrios Colevas, Ann Setser, Valerie Rusch, David Jaques, Volker Budach, Corey Langer, Barbara Murphy, Richard Cumberlin, C. Norman Coleman, and Philip Rubin. CTCAE v3.0: development of a comprehensive grading system for the adverse effects of cancer treatment. *Seminars in Radiation Oncology*, 13(3):176–181, 2003. ISSN 1053–4296. doi: 10.1016/S1053-4296(03)00031-6. URL http://dx.doi.org/10.1016/S1053-4296(03)00031-6. (Cited on page 75.)

[278] S Dische, MF Warburton, D Jones, and E Lartigau. The recording of morbidity related to radiotherapy. *Radiotherapy and Oncology*, 16(2): 103–108, 1989. (Cited on page 75.)

[279] Maria Goleń, Krzysztof Składowski, Andrzej Wygoda, Wiesława Przeorek, Bolesław Pilecki, Mariusz Syguła, Bogusław Maciejewski, and Zofia Kołosza. A comparison of two scoring systems for late radiation toxicity in patients after radiotherapy for head and neck cancer. *Reports of Practical Oncology & Radiotherapy*, 10(4):179–192, 2005. (Cited on pages 75 and 76.)

[280] James D Cox, JoAnn Stetz, and Thomas F Pajak. Toxicity criteria of the radiation therapy oncology group (RTOG) and the european organization for research and treatment of cancer (EORTC). *International Journal of Radiation Oncology Biology Physics*, 31(5): 1341–1346, 1995. (Cited on page 75.)

[281] Klaudia U Hunter, Matthew Schipper, Felix Y Feng, Teresa Lyden, Mark Haxer, Carol-Anne Murdoch-Kinch, Benjamin Cornwall,

Connie SY Lee, Douglas B Chepeha, and Avraham Eisbruch. Toxicities affecting quality of life after chemo-IMRT of oropharyngeal cancer: prospective study of patient-reported, observer-rated, and objective outcomes. *International Journal of Radiation Oncology* Biology* Physics*, 85(4):935–940, 2013. (Cited on page 75.)

[282] Ulrike Hoeller, Silke Tribius, Antje Kuhlmey, Kai Grader, Fabian Fehlauer, and Winfried Alberti. Increasing the rate of late toxicity by changing the score? a comparison of RTOG/EORTC and LENT/SOMA scores. *International Journal of Radiation Oncology Biology Physics*, 55(4):1013–1018, 2003. (Cited on pages 75 and 76.)

[283] Hanna R Mortensen, Kenneth Jensen, Karin Aksglæde, Marie Behrens, and Cai Grau. Late dysphagia after IMRT for head and neck cancer and correlation with dose–volume parameters. *Radiotherapy and Oncology*, 107(3):288–294, 2013. (Cited on pages 75 and 101.)

[284] Fabrice Denis, Pascal Garaud, Etienne Bardet, Marc Alfonsi, Christian Sire, Thierry Germain, Philippe Bergerot, Béatrix Rhein, Jacques Tortochaux, Patrick Oudinot, et al. Late toxicity results of the GORTEC 94-01 randomized trial comparing radiotherapy with concomitant radiochemotherapy for advanced-stage oropharynx carcinoma: comparison of LENT-SOMA, RTOG-EORTC, and NCI-CTC scoring systems. *International Journal of Radiation Oncology Biology Physics*, 55(1):93–98, 2003. (Cited on page 76.)

[285] Søren M Bentzen, Louis S Constine, Joseph O Deasy, Avi Eisbruch, Andrew Jackson, Lawrence B Marks, Randall K Ten Haken, and Ellen D Yorke. Quantitative analyses of normal tissue effects in the clinic (QUANTEC): an introduction to the scientific issues. *International Journal of Radiation Oncology Biology Physics*, 76(3): S3–S9, 2010. (Cited on pages 76, 76, and 171.)

[286] AR Hand and Betty Ho. Liquid-diet-induced alterations of rat parotid acinar cells studied by electron microscopy and enzyme cytochemistry. *Archives of oral biology*, 26(5):369–380, 1981. (Cited on page 76.)

[287] Walter H Wilborn and Charlotte A Schneyer. Ultrastructural changes of rat parotid glands induced by a diet of liquid metrecal. *Zeitschrift für Zellforschung und Mikroskopische Anatomie*, 103(1):1–11, 1970. (Cited on pages 76 and 76.)

[288] Naveen Chhabra, Shruti Chhabra, and Monica Dewan. Post traumatic parotid fistula: A case report. *Journal of Maxillofacial & Oral Surgery*, 7(2), 2008. (Cited on page 77.)

[289] JP Teare, C Spedding, MW Whitehead, SM Greenfield, SJ Challacombe, and RPH Thompson. Omeprazole and dry mouth. *Scandinavian journal of gastroenterology*, 30(3):216–218, 1995. (Cited on pages 77 and 77.)

[290] CBE Scully et al. Drug effects on salivary glands: dry mouth. *Oral diseases*, 9(4):165–176, 2003. (Cited on page 77.)

[291] Vidhi Vinayak, Rajeshwari G Annigeri, Hashikesh A Patel, Sachin Mittal, et al. Adverse affects of drugs on saliva and salivary glands. *Journal of Orofacial Sciences*, 5(1):15, 2013. (Cited on page 77.)

[292] K Mossman, A Shatzman, and J Chencharick. Long-term effects of radiotherapy on taste and salivary function in man. *International Journal of Radiation Oncology Biology Physics*, 8(6):991–997, 1982. (Cited on page 77.)

[293] C-M Eneroth, CO Henrikson, and PÅ Jakobsson. The effect of irradiation in high doses on parotid glands. *Acta otolaryngologica*, 71 (1-6):349–356, 1971. (Cited on page 78.)

[294] H-J Guchelaar, A Vermes, and JH Meerwaldt. Radiation-induced xerostomia: pathophysiology, clinical course and supportive treatment. *Supportive care in cancer*, 5(4):281–288, 1997. (Cited on pages 78, 78, and 79.)

[295] Graydon Charles Smith. Salivary gland dysfunction and xerostomia. Master's thesis, University of Sydney, 1980. (Cited on pages 78 and 78.)

[296] Veerle Persy and Patrick DâĂŹHaese. Vascular calcification and bone disease: the calcification paradox. *Trends in molecular medicine*, 15(9): 405–416, 2009. (Cited on page 78.)

[297] GH Carpenter and E Cotroneo. Salivary gland regeneration. In AS Tucker and I Miletich, editors, *Salivary Glands: Development, Adaptations, and Disease*, chapter 7, pages 107–128. Karger Publishers, London, 2010. (Cited on pages 78, 79, 80, 80, and 80.)

[298] LC Stephens, GK King, LJ Peters, K Kian Ang, TE Schultheiss, and JH Jardine. Acute and late radiation injury in rhesus monkey parotid glands. evidence of interphase cell death. *The American journal of pathology*, 124(3):469, 1986. (Cited on pages 79 and 79.)

[299] O Grundmann, GC Mitchell, and KH Limesand. Sensitivity of salivary glands to radiation: from animal models to therapies. *Journal of dental research*, 88(10):894–903, 2009. (Cited on page 79.)

[300] Rob P Coppes, Ad F Roffel, Liekele JW Zeilstra, Arjan Vissink, and Antonius WT Konings. Early radiation effects on muscarinic receptor-induced secretory responsiveness of the parotid gland in the freely moving rat. *Radiation research*, 153(3):339–346, 2000. (Cited on page 79.)

[301] RP Coppes, A Meter, SP Latumalea, AF Roffel, and HH Kampinga. Defects in muscarinic receptor-coupled signal transduction in isolated parotid gland cells after in vivo irradiation: evidence for a non-DNA target of radiation. *British journal of cancer*, 92(3):539–546, 2005. (Cited on page 79.)

[302] RP Coppes and MA Stokman. Stem cells and the repair of radiation-induced salivary gland damage. *Oral diseases*, 17(2):143–153, 2011. (Cited on pages 79, 80, 80, 80, and 102.)

[303] Antonius WT Konings, Rob P Coppes, and Arjan Vissink. On the mechanism of salivary gland radiosensitivity. *International Journal of Radiation Oncology* Biology* Physics*, 62(4):1187–1194, 2005. (Cited on pages 79 and 80.)

[304] B Nabaa, K Takahashi, T Sasaki, A Okizaki, and T Aburano. Assessment of salivary gland dysfunction after radioiodine therapy for thyroid carcinoma using non-contrast-enhanced CT: the significance of changes in volume and attenuation of the glands. *American J. Neuroradiol.*, 33(10):1964–1970, 2012. (Cited on page 79.)

[305] Keiko Teshima, Ryuji Murakami, Etsuji Tomitaka, Tomoko Nomura, Ryo Toya, Akimitsu Hiraki, Hideki Nakayama, Toshinori Hirai, Masanori Shinohara, Natsuo Oya, et al. Radiation-induced parotid gland changes in oral cancer patients: correlation between parotid volume and saliva production. *Japanese Journal of Clinical Oncology*, 40(1):42–46, 2010. (Cited on pages 79, 104, and 169.)

[306] Leigh C Anderson, Ahmed H Suleiman, and John R Garrett. Morphological effects of diabetes on the granular ducts and acini of the rat submandibular gland. *Microscopy research and technique*, 27(1): 61–70, 1994. (Cited on page 79.)

[307] J Scott, P Liu, and PM Smith. Morphological and functional characteristics of acinar atrophy and recovery in the duct-ligated parotid gland of the rat. *Journal of dental research*, 78(11):1711–1719, 1999. (Cited on pages 80 and 80.)

[308] Shigeru Takahashi, Estelle Schoch, and Neal I Walker. Origin of acinar cell regeneration after atrophy of the rat parotid induced by duct obstruction. *International journal of experimental pathology*, 79(5): 293–301, 1998. (Cited on page 80.)

[309] S Takahashi, K Shinzato, S Nakamura, T Domon, T Yamamoto, and M Wakita. Cell death and cell proliferation in the regeneration of atrophied rat submandibular glands after duct ligation. *Journal of oral pathology & medicine*, 33(1):23–29, 2004. (Cited on page 80.)

[310] S Takahashi, S Nakamura, T Domon, T Yamamoto, and M Wakita. Active participation of apoptosis and mitosis in sublingual gland regeneration of the rat following release from duct ligation. *Journal of molecular histology*, 36(3):199–205, 2005. (Cited on page 80.)

[311] Samira M Osailan, Gordon B Proctor, Guy H Carpenter, Katherine L Paterson, and Mark McGurk. Recovery of rat submandibular salivary gland function following removal of obstruction: a sialometrical and sialochemical study. *International journal of experimental pathology*, 87 (6):411–423, 2006. (Cited on pages 80 and 80.)

[312] GH Carpenter, Nina Khosravani, Jörgen Ekström, SM Osailan, KP Paterson, and GB Proctor. Altered plasticity of the parasympathetic innervation in the recovering rat submandibular gland following extensive atrophy. *Experimental physiology*, 94(2): 213–219, 2009. (Cited on page 80.)

[313] Carolina de la Cal, Alejandro Lomniczi, Claudia E Mohn, Andrea De Laurentiis, Mariana Casal, Ana Chiarenza, Dante Paz, Samuel M McCann, Valeria Rettori, and Juan Carlos Elverdin. Decrease in salivary secretion by radiation mediated by nitric oxide and prostaglandins. *Neuroimmunomodulation*, 13(1):19–27, 2006. (Cited on page 80.)

[314] RP Coppes, LJW Zeilstra, HH Kampinga, and AWT Konings. Early to late sparing of radiation damage to the parotid gland by adrenergic and muscarinic receptor agonists. *British journal of cancer*, 85(7):1055, 2001. (Cited on page 80.)

[315] RM Nagler. The enigmatic mechanism of irradiation-induced damage to the major salivary glands. *Oral diseases*, 8(3):141–146, 2002. (Cited on page 80.)

[316] Everett E Vokes, Ralph R Weichselbaum, Scott M Lippman, and Waun Ki Hong. Head and neck cancer. *New England Journal of Medicine*, 328(3):184–194, 1993. (Cited on page 81.)

[317] Amichay Meirovitz, Carol Anne Murdoch-Kinch, Mathew Schipper, Charlie Pan, and Avraham Eisbruch. Grading xerostomia by physicians or by patients after intensity-modulated radiotherapy of head-and-neck cancer. *International Journal of Radiation Oncology* Biology* Physics*, 66(2):445–453, 2006. (Cited on page 82.)

[318] MK Kam, SF Leung, B Zee, PHK Choi, MRC Chau, KY Cheung, JJS Suen, PML Teo, WH Kwan, and ATC Chan. Impact of intensity-modulated radiotherapy (IMRT) on salivary gland function in early-stage nasopharyngeal carcinoma (NPC) patients: A prospective randomized study. *J Clin Oncol*, 23(Suppl 1):5501, 2005. (Cited on page 82.)

[319] Fred R Burlage, Justin Pijpe, Rob P Coppes, Martin EW Hemels, Harm Meertens, Alof Canrinus, and Arjan Vissink. Variability of flow rate when collecting stimulated human parotid saliva. *European Journal of Oral Sciences*, 113(5):386–390, 2005. (Cited on pages 85 and 86.)

[320] M Bergdahl and J Bergdahl. Low unstimulated salivary flow and subjective oral dryness: association with medication, anxiety, depression, and stress. *Journal of Dental Research*, 79(9):1652–1658, 2000. (Cited on page 85.)

[321] Dena Fischer and Jonathon A Ship. Effect of age on variability of parotid salivary gland flow rates over time. *Age and Ageing*, 28(6): 557–561, 1999. (Cited on page 85.)

[322] Colin Dawes. Effects of diet on salivary secretion and composition. *Journal of dental research*, 49(6):1263–1272, 1970. (Cited on pages 86 and 110.)

[323] J.R. Taylor. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. A series of books in physics. University Science Books, 1997. ISBN 9780935702750. URL https://books.google.ca/books?id=giFQcZub80oC. (Cited on pages 86, 91, and 91.)

[324] Haley Clark, Allan Hovan, Vitali Moiseenko, Steven Thomas, Jonn Wu, and Stefan Reinsberg. Regional radiation dose susceptibility within the parotid gland: Effects on salivary loss and recovery. *Medical Physics*, 42(4):2064–2071, 2015. doi: http://dx.doi.org/10.1118/1.4915077. URL http://scitation.aip.org/content/aapm/journal/medphys/42/4/10.1118/1.4915077. (Cited on pages 88, 140, 145, 146, 149, 171, 172, and 202.)

[325] Lidia Strigari, Marcello Benassi, Giorgio Arcangeli, Vicente Bruzzaniti, Giuseppe Giovinazzo, and Laura Marucci. A novel dose constraint to reduce xerostomia in head-and-neck cancer patients treated with intensity-modulated radiotherapy. *International Journal of Radiation Oncology* Biology* Physics*, 77(1):269–276, 2010. (Cited on page 88.)

[326] Piet Dirix and Sandra Nuyts. Evidence-based organ-sparing radiotherapy in head and neck cancer. *The lancet oncology*, 11(1): 85–91, 2010. (Cited on pages 88 and 89.)

[327] Najat MA Farsi. Signs of oral dryness in relation to salivary flow rate, ph, buffering capacity and dry mouth complaints. *BMC Oral Health*, 7 (1):1, 2007. (Cited on page 88.)

[328] Changyong Feng, Hongyue Wang, Naiji Lu, Tian Chen, Hua He, Ying Lu, and Xin M Tu. Log-transformation and its implications for data analysis. *Shanghai archives of psychiatry*, 26(2):105–109, 2014. (Cited on page 92.)

[329] Michael Greenacre. Data reporting and visualization in ecology. *Polar Biology*, 39(11):2189–2205, 2016. (Cited on page 92.)

[330] Marco Anselmino, Giovanni Zaninotto, Mario Costantini, Pierantonio Ostuni, Aurora Ianniello, Chiara Boccu, Andrea Doria, Silvano Todesco, and Ermanno Ancona. Esophageal motor function in primary

sjögren's syndrome correlation with dysphagia and xerostomia. *Digestive diseases and sciences*, 42(1):113–118, 1997. (Cited on page 95.)

[331] Bukola Ojo, Eric M Genden, Marita S Teng, Kathrin Milbury, Krzysztof J Misiukiewicz, and Hoda Badr. A systematic review of head and neck cancer quality of life assessment instruments. *Oral oncology*, 48(10):923–937, 2012. (Cited on page 95.)

[332] Amy Y Chen, Ralph Frankowski, Julie Bishop-Leone, Tiffany Hebert, Stacy Leyk, Jan Lewin, and Helmuth Goepfert. The development and validation of a dysphagia-specific quality-of-life questionnaire for patients with head and neck cancer: the MD anderson dysphagia inventory. *Archives of Otolaryngology–Head & Neck Surgery*, 127(7): 870–876, 2001. (Cited on page 96.)

[333] Gary D Slade. Derivation and validation of a short-form oral health impact profile. *Community dentistry and oral epidemiology*, 25(4): 284–290, 1997. (Cited on page 96.)

[334] M Navazesh, C Christensen, and V Brightman. Clinical criteria for the diagnosis of salivary gland hypofunction. *Journal of Dental Research*, 71(7):1363–1369, 1992. (Cited on pages 96 and 98.)

[335] Judith M Roesink, Maria Schipper, Wim Busschers, Cornelis PJ Raaijmakers, and Chris HJ Terhaard. A comparison of mean parotid gland dose with measures of parotid gland function after radiotherapy for head-and-neck cancer: Implications for future trials. *International Journal of Radiation Oncology Biology Physics*, 63(4):1006–1009, 2005. (Cited on page 96.)

[336] AJ Carlson and AL Crittenden. The relation of ptyalin concentration to the diet and to the rate of secretion of the saliva. *American Journal of Physiology–Legacy Content*, 26(1):169–177, 1910. (Cited on page 97.)

[337] KS Lashley. Reflex secretion of the human parotid gland. *Journal of Experimental Psychology*, 1(6):461, 1916. (Cited on page 97.)

[338] A Vissink, AWT Konings, EE Ligeon, et al. An adaptation of the lashley cup for use in rat saliva collection. *Archives of oral biology*, 34 (7):577–578, 1989. (Cited on page 97.)

[339] Leon H Schneyer. Method for the collection of separate submaxillary and sublingual salivas in man. *Journal of dental research*, 34(2): 257–261, 1955. (Cited on page 97.)

[340] CA Schneyer and LH Schneyer. Electrolyte levels of rat salivary secretions in relation to fluid-flow rate. *American Journal of Physiology–Legacy Content*, 199(1):55–58, 1960. (Cited on page 97.)

[341] M Navazesh and CM Christensen. A comparison of whole mouth resting and stimulated salivary measurement procedures. *Journal of Dental Research*, 61(10):1158–1162, 1982. (Cited on page 97.)

[342] Andy Wolff, Mel Rosenberg, et al. A simple technique for the determination of salivary gland hypofunction. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, 94(2): 175–178, 2002. (Cited on page 98.)

[343] Philip P Connell and Samuel Hellman. Advances in radiotherapy and implications for the next century: a historical perspective. *Cancer research*, 69(2):383–392, 2009. (Cited on page 99.)

[344] Anders Brahme. Dosimetric precision requirements in radiation therapy. *Acta Radiologica: Oncology*, 23(5):379–391, 1984. (Cited on page 99.)

[345] Qiuwen Wu, Radhe Mohan, Andrzej Niemierko, and Rupert Schmidt-Ullrich. Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose. *International Journal of Radiation Oncology* Biology* Physics*, 52(1):224–235, 2002. (Cited on page 99.)

[346] Issam El Naqa, Jeffrey Bradley, Angel I Blanco, Patricia E Lindsay, Milos Vicic, Andrew Hope, and Joseph O Deasy. Multivariable modeling of radiotherapy outcomes, including dose–volume and clinical factors. *International Journal of Radiation Oncology Biology Physics*, 64(4):1275–1286, 2006. (Cited on pages 100, 100, 104, and 104.)

[347] Antonetta C Houweling, Marielle EP Philippens, Tim Dijkema, Judith M Roesink, Chris HJ Terhaard, Cornelis Schilstra, Randall K Ten Haken, Avraham Eisbruch, and Cornelis PJ Raaijmakers. A comparison of dose–response models for the parotid gland in a large group of head-and-neck cancer patients. *International Journal of*

Radiation Oncology Biology Physics, 76(4):1259–1265, 2010. (Cited on pages 100 and 100.)

[348] Antonius WT Konings, Femmy Cotteleer, Hette Faber, Peter van Luijk, Harm Meertens, and Rob P Coppes. Volume effects and region-dependent radiosensitivity of the parotid gland. *International Journal of Radiation Oncology Biology Physics*, 62(4):1090–1095, 2005. (Cited on pages 100, 101, 126, 145, 168, 172, and 215.)

[349] Antonius WT Konings, Hette Faber, Femmy Cotteleer, Arjan Vissink, and Rob P Coppes. Secondary radiation damage as the main cause for unexpected volume effects: a histopathologic study of the parotid gland. *IJROBP*, 64(1):98–105, 2006. (Cited on pages 100 and 215.)

[350] Anthony B Wolbarst. Optimization of radiation therapy II: The critical-voxel model. *International Journal of Radiation Oncology* Biology* Physics*, 10(5):741–745, 1984. (Cited on page 100.)

[351] P Källman, A Ågren, and A Brahme. Tumour and normal tissue responses to fractionated non-uniform dose delivery. *International journal of radiation biology*, 62(2):249–262, 1992. (Cited on page 100.)

[352] Joseph O Deasy and Issam El Naqa. Image-based modeling of normal tissue complication probability for radiation therapy. In *Radiation Oncology Advances*, pages 211–252. Springer, 2008. (Cited on page 100.)

[353] John William Hopewell and Klaus-Rüdiger Trott. Volume effects in radiobiology as applied to radiotherapy. *Radiotherapy and Oncology*, 56(3):283–288, 2000. (Cited on pages 100 and 186.)

[354] Peter van Luijk, Hendrik P Bijl, Antonius WT Konings, Albert J Kogel, and Jacobus M Schippers. Data on dose–volume effects in the rat spinal cord do not support existing NTCP models. *International Journal of Radiation Oncology Biology Physics*, 61(3):892–900, 2005. (Cited on page 101.)

[355] Tim Dijkema, Chris HJ Terhaard, Judith M Roesink, Pètra M Braam, Carla H van Gils, Marinus A Moerland, and Cornelis PJ Raaijmakers. Large cohort dose–volume response analysis of parotid gland function after radiotherapy: intensity-modulated versus conventional radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 72(4):1101–1109, 2008. (Cited on pages 101 and 146.)

[356] Peter van Luijk, Hette Faber, Jacobus M Schippers, Sytze Brandenburg, Johannes A Langendijk, Harm Meertens, and Robert P Coppes. Bath and shower effects in the rat parotid gland explain increased relative risk of parotid gland dysfunction after intensity-modulated radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 74(4):1002–1005, 2009. (Cited on pages 101, 103, 146, 168, 171, 171, 202, 215, and 215.)

[357] Peter van Luijk, Sarah Pringle, Joseph O Deasy, Vitali V Moiseenko, Hette Faber, Allan Hovan, Mirjam Baanstra, Hans P van der Laan, Roel GJ Kierkels, Arjen van der Schaaf, et al. Sparing the region of the salivary gland containing stem cells preserves saliva production after radiotherapy for head and neck cancer. *Science translational medicine*, 7(305):305ra147–305ra147, 2015. (Cited on pages 102, 102, 126, 127, 128, 128, 140, 145, 147, 168, 171, 172, 186, and 214.)

[358] AB Miah, SL Gulliford, J Morden, KL Newbold, SA Bhide, SH Zaidi, E Hall, KJ Harrington, and CM Nutting. Recovery of salivary function: Contralateral parotid-sparing intensity-modulated radiotherapy versus bilateral superficial lobe parotid-sparing intensity-modulated radiotherapy. *Clinical Oncology*, 2016. (Cited on pages 102, 126, 140, 145, 147, 168, 171, and 172.)

[359] Emanuele Cotroneo, Gordon B Proctor, Katherine L Paterson, and Guy H Carpenter. Early markers of regeneration following ductal ligation in rat submandibular gland. *Cell and tissue research*, 332(2): 227–235, 2008. (Cited on page 102.)

[360] Isabelle MA Lombaert, Jeanette F Brunsting, Pieter K Wierenga, Hette Faber, Monique A Stokman, Tineke Kok, Willy H Visser, Harm H Kampinga, Gerald de Haan, and Robert P Coppes. Rescue of salivary gland function after stem cell transplantation in irradiated glands. *PloS one*, 3(4):e2063, 2008. (Cited on pages 103 and 146.)

[361] Yoshinori Sumita, Younan Liu, Saeed Khalili, Ola M Maria, Dengsheng Xia, Sharon Key, Ana P Cotrim, Eva Mezey, and Simon D Tran. Bone marrow-derived cells rescue salivary gland function in mice with head and neck irradiation. *The International Journal of Biochemistry & Cell Biology*, 43(1):80–87, 2011. (Cited on pages 103 and 146.)

[362] Lalitha SY Nanduri, Martti Maimets, Sarah A Pringle, Marianne van der Zwaag, Ronald P van Os, and Robert P Coppes. Regeneration

of irradiated salivary glands with stem cell marker expressing cells. *Radiotherapy and Oncology*, 99(3):367–372, 2011. (Cited on page 103.)

[363] Ayumi Sato, Kenji Okumura, Shirou Matsumoto, Kiyoko Hattori, Shinzaburo Hattori, Masanori Shinohara, and Fumio Endo. Isolation, tissue localization, and cellular characterization of progenitors derived from adult human salivary glands. *Cloning and stem cells*, 9(2): 191–205, 2007. (Cited on page 103.)

[364] Erwin Gorjup, Sandra Danner, Nicole Rotter, Jens Habermann, Ute Brassat, Tim H Brummendorf, Sascha Wien, Andreas Meyerhans, Barbara Wollenberg, Charli Kruse, et al. Glandular tissue from human pancreas and salivary gland yields similar stem cell populations. *European journal of cell biology*, 88(7):409–421, 2009. (Cited on page 103.)

[365] D Andreadis, A Bakopoulou, G Leyhausen, A Epivatianos, J Volk, A Markopoulos, and W Geurtsen. Minor salivary glands of the lips: a novel, easily accessible source of potential stem/progenitor cells. *Clinical oral investigations*, 18(3):847–856, 2014. (Cited on page 103.)

[366] Sarah Pringle, Ronald Van Os, and Robert P Coppes. Concise review: Adult salivary gland stem cells and a potential therapy for xerostomia. *Stem cells*, 31(4):613–619, 2013. (Cited on page 103.)

[367] Sarah Pringle, Martti Maimets, Marianne van der Zwaag, Monique A Stokman, Djoke van Gosliga, Erik Zwart, Max JH Witjes, Gerald de Haan, Ronald van Os, and Rob P Coppes. Human salivary gland stem cells functionally restore radiation damaged salivary glands. *STEM CELLS*, 34(3):640–652, 2016. (Cited on page 103.)

[368] Padma Pradeepa Srinivasan, Vaishali N Patel, Shuang Liu, Daniel A Harrington, Matthew P Hoffman, Xinqiao Jia, Robert L Witt, Mary C Farach-Carson, and Swati Pradhan-Bhatt. Primary salivary human stem/progenitor cells undergo microenvironment-driven acinar-like differentiation in hyaluronate hydrogel culture. *Stem Cells Translational Medicine*, pages sctm–2016, 2016. (Cited on page 103.)

[369] Bing Liu, Michael R Dion, M Marianne Jurasic, Gretchen Gibson, and Judith A Jones. Xerostomia and salivary hypofunction in vulnerable elders: prevalence and etiology. *Oral surgery, oral medicine, oral pathology and oral radiology*, 114(1):52–60, 2012. (Cited on page 104.)

[370] AE Kossioni and HC Karkazis. Socio-medical condition and oral functional status in an older institutionalised population. *Gerodontology*, 16(1):21–28, 1999. (Cited on page 104.)

[371] Fu-Min Fang, Wen-Ling Tsai, Chih-Yen Chien, Herng-Chia Chiu, and Chong-Jong Wang. Health-related quality of life outcome for oral cancer survivors after surgery and postoperative radiotherapy. *Japanese journal of clinical oncology*, 34(11):641–646, 2004. (Cited on page 104.)

[372] Kenneth Jensen, Anders Bonde Jensen, and Cai Grau. Smoking has a negative impact upon health related quality of life after treatment for head and neck cancer. *Oral oncology*, 43(2):187–192, 2007. (Cited on page 104.)

[373] Sonia A Duffy, Jeffrey E Terrell, Marcia Valenstein, David L Ronis, Laurel A Copeland, and Mary Connors. Effect of smoking, alcohol, and depression on the quality of life of head and neck cancer patients. *General hospital psychiatry*, 24(3):140–147, 2002. (Cited on page 104.)

[374] Judith M. Roesink, Marinus A. Moerland, Jan J. Battermann, Gerrit Jan Hordijk, and Chris H.J. Terhaard. Quantitative dose-volume response analysis of changes in parotid gland function after radiotheraphy in the head-and-neck region. *IJROBP*, 51(4): 938–946, 2001. ISSN 0360-3016. (Cited on pages 105, 172, 185, and 187.)

[375] Tsair-Fwu Lee, Ming-Hsiang Liou, Yu-Jie Huang, Pei-Ju Chao, Hui-Min Ting, Hsiao-Yi Lee, and Fu-Min Fang. LASSO NTCP predictors for the incidence of xerostomia in patients with head and neck squamous cell carcinoma and nasopharyngeal carcinoma. *Scientific reports*, 4, 2014. (Cited on page 105.)

[376] Mark Barnes, Manpreet S Tiwana, Stacy Miller, Andrew Kiraly, Ivo A Olivotto, Scott Emmons, and Robert A Olson. Palliative radiotherapy fractionation schedules prescribed are dependent on the distance a patient travels to receive treatment. *Radiotherapy and Oncology*, 117 (2):390–392, 2015. (Cited on page 105.)

[377] Sheila Payne, Nicola Jarrett, and D Jeffs. The impact of travel on cancer patientsâĂŹ experiences of treatment: a literature review. *European journal of cancer care*, 9(4):197–203, 2000. (Cited on pages 105, 106, and 107.)

[378] Robert A Olson, Alan Nichol, Nadine R Caron, Ivo A Olivotto, Caroline Speers, Stephen Chia, Ashley Davidson, Andy Coldman, Chris Bajdik, and Scott Tyldesley. Effect of community population size on breast cancer screening, stage distribution, treatment use and outcomes. *Canadian Journal of Public Health/Revue Canadienne de Sante'e Publique*, pages 46–52, 2012. (Cited on page 105.)

[379] GKD Crozier and Françoise Baylis. The ethical physician encounters international medical travel. *Journal of Medical Ethics*, 36(5):297–301, 2010. (Cited on page 105.)

[380] Michael D Horowitz, Jeffrey A Rosensweig, and Christopher A Jones. Medical tourism: globalization of the healthcare marketplace. *Medscape General Medicine*, 9(4):33, 2007. (Cited on page 105.)

[381] Leigh Turner. CanadaâĂŹs turbulent medical tourism industry. *Canadian Family Physician*, 58(4):371–373, 2012. (Cited on page 106.)

[382] Josef Woodman. *Patients beyond borders: Everybody's guide to affordable, world-class medical travel*. Healthy Travel Media, 2009. (Cited on page 106.)

[383] Rory Johnston, Valorie A Crooks, Jeremy Snyder, and Shafik Dharamsi. Canadian family doctorsâĂŹ roles and responsibilities toward outbound medical tourists âĂIJour true role is... within the confines of our systemâĂİ. *Canadian Family Physician*, 59(12): 1314–1319, 2013. (Cited on page 106.)

[384] Jeremy Snyder, Valorie A Crooks, Rory Johnston, and Paul Kingsbury. What do we know about canadian involvement in medical tourism? a scoping review. *Open Medicine*, 5(3):139–148, 2011. (Cited on page 106.)

[385] Edward H Simpson. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 238–241, 1951. (Cited on pages 106 and 168.)

[386] J Mark Stephens, Xiaoyan Li, Maureen Reiner, and Spiros Tzivelekis. Annual patient and caregiver burden of oncology clinic visits for granulocyte-colony stimulating factor therapy in the US. *Journal of medical economics*, 19(5):537–547, 2016. (Cited on pages 107 and 107.)

[387] Margaret I Fitch, Ross E Gray, Tom McGowan, Ian Brunskill, Shawn Steggles, Scott Sellick, Andrea Bezjak, and Donna McLeese. Travelling for radiation cancer treatment: patient perspectives. *Psycho-Oncology*, 12(7):664–674, 2003. (Cited on page 107.)

[388] Wendy S Schain. Barriers to clinical trials: Part II: Knowledge and attitudes of potential participants. *Cancer*, 74(S9):2666–2671, 1994. (Cited on pages 107 and 108.)

[389] Christine L Paul, Alix E Hall, Mariko L Carey, Emilie C Cameron, and Tara Clinton-McHarg. Access to care and impacts of cancer on daily life: do they differ for metropolitan versus regional hematological cancer survivors? *The Journal of Rural Health*, 29(s1):s43–s50, 2013. (Cited on pages 107, 108, and 108.)

[390] Bruce E Johnson, Daniel C Ihde, Paul A Bunn, Bruce Becker, Thomas Walsh, Zelig R Weinstein, Mary J Matthews, Jacqueline Whang-Peng, Robert W Makuch, Anita Johnston-Early, et al. Patients with small-cell lung cancer treated with combination chemotherapy with or without irradiation: data on potential cures, chronic toxicities, and late relapses after a five-to eleven-year follow-up. *Annals of internal medicine*, 103(3):430–438, 1985. (Cited on page 107.)

[391] Wayne M Koch, Miriam Lango, Duane Sewell, Marianna Zahurak, and David Sidransky. Head and neck cancer in nonsmokers: A distinct clinical and molecular entity. *The Laryngoscope*, 109(10):1544–1551, 1999. (Cited on pages 107 and 107.)

[392] B Bergman and S Sorenson. Smoking and effect of chemotherapy in small cell lung cancer. *European Respiratory Journal*, 1(10):932–937, 1988. (Cited on page 107.)

[393] George P Browman, Gene Wong, Ian Hodson, Jinka Sathya, Rosemary Russell, Lisa McAlpine, Peter Skingley, and Mark N Levine. Influence of cigarette smoking on the efficacy of radiation therapy in head and neck cancer. *New England Journal of Medicine*, 328(3):159–163, 1993. (Cited on page 107.)

[394] André Fortin, Chang Shu Wang, and Éric Vigneault. Influence of smoking and alcohol drinking behaviors on treatment outcomes of patients with squamous cell carcinomas of the head and neck. *International Journal of Radiation Oncology* Biology* Physics*, 74(4): 1062–1069, 2009. (Cited on page 107.)

[395] Fadlo R Khuri, Edward S Kim, J Jack Lee, Rodger J Winn, Steven E Benner, Scott M Lippman, Karen K Fu, Jay S Cooper, Everett E Vokes, Robert M Chamberlain, et al. The impact of smoking status, disease stage, and index tumor site on second primary tumor incidence and tumor recurrence in the head and neck retinoid chemoprevention trial. *Cancer Epidemiology Biomarkers & Prevention*, 10(8):823–829, 2001. (Cited on page 107.)

[396] Robert A. Olson. Population-based assessment of relationship between volume of practice and outcomes in head and neck cancer patients. Presentation at the BCCA Provincial Head and Neck Tumor Group Retreat 25 November 2016, 2016. (Cited on page 108.)

[397] Raymond K Tsang and William I Wei. Salvage surgery for nasopharyngeal cancer. *World Journal of Otorhinolaryngology-Head and Neck Surgery*, 1(1):34–43, 2015. (Cited on page 109.)

[398] William L Wei, Kam H Lam, and Jonathan ST Sham. New approach to the nasopharynx: the maxillary swing approach. *Head & neck*, 13 (3):200–207, 1991. (Cited on page 109.)

[399] William I Wei and Dora LW Kwong. Current management strategy of nasopharyngeal carcinoma. *Clinical and experimental otorhinolaryngology*, 3(1):1–12, 2010. (Cited on page 109.)

[400] C Yu Mimi and Jian-Min Yuan. Epidemiology of nasopharyngeal carcinoma. In *Seminars in cancer biology*, volume 12, pages 421–429. Elsevier, 2002. (Cited on page 109.)

[401] DE Peterson. Pretreatment strategies for infection prevention in chemotherapy patients. *NCI monographs: a publication of the National Cancer Institute*, (9):61–71, 1989. (Cited on page 109.)

[402] Michael D Turner and Jonathan A Ship. Dry mouth and its effects on the oral health of elderly people. *The Journal of the American Dental Association*, 138:S15–S20, 2007. (Cited on page 109.)

[403] Jean-Pierre Pignon, Aurelie Le Maitre, Emilie Maillard, Jean Bourhis, et al. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17,346 patients. *Radiotherapy and Oncology*, 92(1):4–14, 2009. (Cited on pages 109 and 110.)

[404] Marianne C Walsh, Lorraine Brennan, J Paul G Malthouse, Helen M Roche, and Michael J Gibney. Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. *The American journal of clinical nutrition*, 84(3):531–539, 2006. (Cited on page 110.)

[405] AJ Munro and S Potter. A quantitative approach to the distress caused by symptoms in patients treated with radical radiotherapy. *British journal of cancer*, 74(4):640, 1996. (Cited on page 110.)

[406] Charlotte Jacobs, DR Goffinet, L Goffinet, M Kohler, WE Fee, et al. Chemotherapy as a substitute for surgery in the treatment of advanced resectable head and neck cancer. *Cancer*, 60(6):1178–83, 1987. (Cited on page 110.)

[407] Adam S Garden, Joshua A Asper, William H Morrison, Naomi R Schechter, Bonnie S Glisson, Merrill S Kies, Jeffrey N Myers, and K Kian Ang. Is concurrent chemoradiation the treatment of choice for all patients with stage III or IV head and neck carcinoma? *Cancer*, 100(6):1171–1178, 2004. (Cited on page 110.)

[408] James D Fearon. Ethnic and cultural diversity by country. *Journal of Economic Growth*, 8(2):195–222, 2003. (Cited on page 110.)

[409] Avraham Eisbruch, Randall K Ten Haken, Hyungjin M Kim, Lon H Marsh, and Jonathan A Ship. Dose, volume, and function relationships in parotid salivary glands following conformal and intensity-modulated irradiation of head and neck cancer. *International Journal of Radiation Oncology* Biology* Physics*, 45(3):577–587, 1999. (Cited on pages 111 and 111.)

[410] Ivo Beetz, Roel JHM Steenbakkers, Olga Chouvalova, Charles R Leemans, Patricia Doornaert, Bernard FAM van der Laan, Miranda EMC Christianen, Arjan Vissink, Henk P Bijl, Peter van Luijk, et al. The QUANTEC criteria for parotid gland dose and their efficacy to prevent moderate to severe patient-rated xerostomia. *Acta Oncologica*, 53(5):597–604, 2014. (Cited on pages 111 and 202.)

[411] H Clark, J Wu, V Moiseenko, R Lee, B Gill, C Duzenli, and S Thomas. Semi-automated contour recognition using DICOMautomaton. *Journal of Physics: Conference Series*, 489(1):012088, 2014. doi: 10.1088/1742-6596/489/1/012088. URL

http://dx.doi.org/10.1088/1742-6596/489/1/012088. http://stacks.iop.org/1742-6596/489/i=1/a=012088. (Cited on pages 116, 131, 148, and 173.)

[412] Haley Clark, Jonn Wu, Vitali Moiseenko, and Steven Thomas. Poster – Thur Eve – 74: Distributed, asynchronous, reactive dosimetric and outcomes analysis using DICOMautomaton. *Medical Physics*, 41(8): 22–22, 2014. ISSN 2473-4209. doi: 10.1118/1.4894935. URL http://dx.doi.org/10.1118/1.4894935. (Cited on page 116.)

[413] Barend Gehrels, Bruno Lalande, and Mateusz Loskot. Generic programming for geometry. BoostCon 2010. (Accessed 25 December 2016.), 2010. URL https://github.com/boostcon/2010_presentations/raw/master/mon/boost_geometry_boostcon2010.pdf. (Cited on page 117.)

[414] The CGAL Project. *CGAL User and Reference Manual*. CGAL Editorial Board, 4.9 edition, 2016. (Cited on pages 117 and 140.)

[415] Sébastien Loriot, Jane Tournois, and Ilker O. Yaz. Polygon mesh processing. In *CGAL User and Reference Manual*. CGAL Editorial Board, 4.9 edition, 2016. (Cited on pages 117 and 140.)

[416] Jack C Morrison. Distance from a point to a line. pages 10–13, 1991. (Cited on page 117.)

[417] Mukesh Prasad. Intersection of line segments. In *Graphics Gems II*, pages 7–9. Academic Press Inc., New York, 1991. (Cited on page 117.)

[418] Ronald N Goldman. Area of planar polygons and volume of polyhedra. In *Graphics gems II*, volume 2, pages 24–28. Academic Press Inc., New York, 1991. (Cited on pages 117 and 119.)

[419] Gerard Bashein and Paul R Detmer. Centroid of a polygon. In *Graphics gems IV*, pages 3–6. Academic Press Professional, Inc., 1994. (Cited on page 118.)

[420] Eric Haines. Point in polygon strategies. In *Graphics gems IV*, pages 24–26. Academic Press Professional, Inc., 1994. (Cited on page 118.)

[421] Ailsa H Land and Alison G Doig. An automatic method of solving discrete programming problems. *Econometrica: Journal of the Econometric Society*, pages 497–520, 1960. (Cited on page 125.)

[422] Jens Clausen. Branch and bound algorithms-principles and examples. *Department of Computer Science, University of Copenhagen*, pages 1–30, 1999. (Cited on page 125.)

[423] Patrenahalli M. Narendra and Keinosuke Fukunaga. A branch and bound algorithm for feature subset selection. *IEEE Transactions on Computers*, 100(9):917–922, 1977. (Cited on page 125.)

[424] Dana S Nau, Vipin Kumar, and Laveen Kanal. General branch and bound, and its relation to A* and AO*. *Artificial Intelligence*, 23(1): 29–58, 1984. (Cited on page 125.)

[425] Tiziana Rancati, Marco Schwarz, Aaron M Allen, Felix Feng, Aron Popovtzer, Bharat Mittal, and Avraham Eisbruch. Radiation dose–volume effects in the larynx and pharynx. *International Journal of Radiation Oncology* Biology* Physics*, 76(3):S64–S69, 2010. (Cited on page 126.)

[426] Laura A Dawson, Brian D Kavanagh, Arnold C Paulino, Shiva K Das, Moyed Miften, X Allen Li, Charlie Pan, Randall K Ten Haken, and Timothy E Schultheiss. Radiation-associated kidney injury. *International Journal of Radiation Oncology* Biology* Physics*, 76(3): S108–S115, 2010. (Cited on page 126.)

[427] Charlie C Pan, Brian D Kavanagh, Laura A Dawson, X Allen Li, Shiva K Das, Moyed Miften, and Randall K Ten Haken. Radiation-associated liver injury. *International Journal of Radiation Oncology* Biology* Physics*, 76(3):S94–S100, 2010. (Cited on page 126.)

[428] Wei Xiao, Zhixiong Lin, Wuzhe Zhang, Mei Li, and Vincent WC Wu. A split-parotid delineation approach for dose optimization in volumetric modulated arc therapy for nasopharyngeal carcinoma patients with parapharyngeal space invasion and level IIa cervical lymph node involvements. *The British journal of radiology*, 89(1060): 20150635, 2016. (Cited on pages 126, 140, 171, and 172.)

[429] Douglas G Bonett. Confidence interval for a coefficient of quartile variation. *Computational Statistics & Data Analysis*, 50(11):2953–2957, 2006. (Cited on page 131.)

[430] Raymond Clare Archibald. *Euclid's Book on Division of Figures: With a Restoration Based on Woepcke's Text and on the Practica Geometriae of Leonardo Pisano*. CUP Archive, 2015. (Cited on page 132.)

[431] Kenneth L Kelly. Twenty-two colors of maximum contrast. *Color Engineering*, 3(26):26–27, 1965. (Cited on page 135.)

[432] Joseph O Deasy, Angel I Blanco, and Vanessa H Clark. CERR: a computational environment for radiotherapy research. *Medical physics*, 30(5):979–985, 2003. (Cited on page 140.)

[433] Csaba Pinter, Andras Lasso, An Wang, David Jaffray, and Gabor Fichtinger. SlicerRT: radiation therapy research toolkit for 3D Slicer. *Medical physics*, 39(10):6332–6338, 2012. (Cited on pages 140 and 143.)

[434] B Boots, A Okabe, and K Sugihara. Spatial tessellations. *Geographical information systems*, 1:503–526, 1999. (Cited on pages 140 and 281.)

[435] W. Fraser. A survey of methods of computing minimax and near-minimax polynomial approximations for functions of a single independent variable. *J. ACM*, 12(3):295–314, 1965. ISSN 0004-5411. doi: 10.1145/321281.321282. (Cited on pages 143 and 280.)

[436] Vladimir Fedorovich Dem'yanov and Vasilij Nikolaevič Malozemov. *Introduction to minimax*. Courier Corporation, 1974. (Cited on pages 143 and 280.)

[437] Bruce J Baum. Radiation-induced salivary hypofunction may become a thing of the past. *Oral Dis.*, 22(2):81–84, 2016. (Cited on page 147.)

[438] Jian Bi. A review of statistical methods for determination of relative importance of correlated predictors and identification of drivers of consumer liking. *J. Sens. Stud.*, 27(2):87–101, 2012. (Cited on pages 147, 151, 151, 164, 166, and 167.)

[439] Kenneth P Burnham and David Anderson. *Model selection and multi-model inference*. Springer, 2 edition, 2003. (Cited on pages 149, 150, 166, and 166.)

[440] Hamparsum Bozdogan. Model selection and akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370, 1987. (Cited on page 150.)

[441] Cort J Willmott and Kenji Matsuura. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1):79–82, 2005. (Cited on pages 150 and 175.)

[442] Harvey J Motulsky and Lennart A Ransnas. Fitting curves to data using nonlinear regression: a practical and nonmathematical review. *The FASEB J.*, 1(5):365–374, 1987. (Cited on pages 150 and 151.)

[443] Richard Harold Lindeman and Richard Harold Lindeman. Introduction to bivariate and multivariate analysis. Technical report, Scott, Foresman, & Co., 1980. (Cited on pages 151 and 165.)

[444] Martin B Wilk and Ram Gnanadesikan. Probability plotting methods for the analysis for the analysis of data. *Biometrika*, 55(1):1–17, 1968. (Cited on page 152.)

[445] Marie Laure Delignette-Muller and Christophe Dutang. fitdistrplus: An R package for fitting distributions. *J. Stat. Soft.*, 64(4):1–34, 2015. (Cited on page 152.)

[446] Robert L Bailey and TR Dell. Quantifying diameter distributions with the weibull function. *Forest Science*, 19(2):97–104, 1973. (Cited on page 153.)

[447] Abraham Charnes, EL Frome, and Po-Lung Yu. The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. *J. Am. Stat. Assoc.*, 71(353):169–171, 1976. (Cited on pages 153 and 164.)

[448] Sara Gustavsson, Björn Fagerberg, Gerd Sallsten, and Eva M Andersson. Regression models for log-normal data: comparing different methods for quantifying the association between abdominal adiposity and biomarkers of inflammation and insulin resistance. *Int J Environ Res Public Health*, 11(4):3521–3539, 2014. (Cited on pages 153 and 164.)

[449] Cécile Ortholan, Emmanuel Chamorey, Karen Benezery, Juliette Thariat, Olivier Dassonville, Gilles Poissonnet, Alexandre Bozec, Philippe Follana, Frédérique Peyrade, Anne Sudaka, et al. Modeling of salivary production recovery after radiotherapy using mixed models: determination of optimal dose constraint for IMRT planning and construction of convenient tools to predict salivary function.

International Journal of Radiation Oncology Biology* Physics*, 73(1): 178–186, 2009. (Cited on pages 171, 172, 184, and 186.)

[450] Zhong-He Wang, Chao Yan, Zhi-Yuan Zhang, Chen-Ping Zhang, Hai-Sheng Hu, Wen-Yong Tu, Jessica Kirwan, and William M Mendenhall. Impact of salivary gland dosimetry on post-IMRT recovery of saliva output and xerostomia grade for head-and-neck cancer patients treated with or without contralateral submandibular gland sparing: a longitudinal study. *International Journal of Radiation Oncology* Biology* Physics*, 81(5):1479–1487, 2011. (Cited on pages 171 and 172.)

[451] HB Zhang, X Lu, SM Huang, L Wang, C Zhao, WX Xia, SW Li, FL Wang, YL Zhu, X Guo, et al. Superficial parotid lobe–sparing delineation approach: a better method of dose optimization to protect the parotid gland in intensity-modulated radiotherapy for nasopharyngeal carcinoma. *Current Oncology*, 20(6):e577, 2013. (Cited on page 172.)

[452] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001. (Cited on pages 174 and 174.)

[453] Andy Liaw and Matthew Wiener. Classification and regression by randomForest. *R news*, 2(3):18–22, 2002. (Cited on page 174.)

[454] André Altmann, Laura Toloşi, Oliver Sander, and Thomas Lengauer. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10):1340–1347, 2010. (Cited on page 174.)

[455] Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinf.*, 8(1):1, 2007. (Cited on pages 174, 174, and 174.)

[456] Torsten Hothorn, Kurt Hornik, and Achim Zeileis. Unbiased recursive partitioning: A conditional inference framework. *JCGS*, 15(3):651–674, 2006. (Cited on page 174.)

[457] Carolin Strobl, James Malley, and Gerhard Tutz. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods*, 14(4):323, 2009. (Cited on pages 174 and 174.)

[458] Carolin Strobl, Torsten Hothorn, and Achim Zeileis. Party on! a new, conditional variable-importance measure for random forests available in the party package. *The R Journal*, 1/2(74.52034):14–17, 2009. (Cited on pages 174 and 187.)

[459] Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974. (Cited on page 175.)

[460] Mervyn Stone. An asymptotic equivalence of choice of model by cross-validation and akaike's criterion. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 44–47, 1977. (Cited on page 175.)

[461] Tianfeng Chai and Roland R Draxler. Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. *Geosci. Model Dev.*, 7(3):1247–1250, 2014. (Cited on page 175.)

[462] Olive Jean Dunn and Virginia Clark. Comparison of tests of equality of dependent correlation coefficients. *J. Am. Stat. Assoc.*, 66(336): 904–908, 1971. (Cited on page 175.)

[463] KR Trott, S Tamou, T Sassy, and Z Kiszel. The effect of irradiated volume on the chronic radiation damage of the rat large bowel. *Strahlentherapie und Onkologie: Organ der Deutschen Rontgengesellschaft*, 171(6):326–331, 1995. (Cited on page 186.)

[464] Siri Beier Jensen, Anne Marie Lynge Pedersen, Arjan Vissink, Elo Andersen, CG Brown, AN Davies, J Dutilh, JS Fulton, L Jankovic, NNF Lopes, et al. A systematic review of salivary gland hypofunction and xerostomia induced by cancer therapies: management strategies and economic impact. *Supportive care in cancer*, 18(8):1061–1079, 2010. (Cited on pages 201 and 212.)

[465] A Vissink, P Luijk, JA Langendijk, and RP Coppes. Current ideas to reduce or salvage radiation damage to salivary glands. *Oral diseases*, 21(1):e1–e10, 2015. (Cited on pages 202 and 202.)

[466] Mike Partridge, Tokihiro Yamamoto, Cai Grau, Morten Høyer, and Ludvig Paul Muren. Imaging of normal lung, liver and parotid gland function for radiotherapy. *Acta Oncologica*, 49(7):997–1011, 2010. (Cited on pages 202, 202, and 202.)

[467] Hans Paul van der Laan, Tara A van de Water, Heleen E van Herpt, Miranda EMC Christianen, Hendrik P Bijl, Erik W Korevaar, Coen R Rasch, Aart A van 'T Veld, Arjen van der Schaaf, Cornelis Schilstra, et al. The potential of intensity-modulated proton radiotherapy to reduce swallowing dysfunction in the treatment of head and neck cancer: A planning comparative study. *Acta oncologica*, 52(3):561–569, 2013. (Cited on page 202.)

[468] Chun-Jung Juan, Cheng-Yu Chen, Yee-Min Jen, Hua-Shan Liu, Yi-Jui Liu, Chun-Jen Hsueh, Chao-Ying Wang, Yu-Ching Chou, Yao-Te Chai, Guo-Shu Huang, et al. Perfusion characteristics of late radiation injury of parotid glands: quantitative evaluation with dynamic contrast-enhanced MRI. *European radiology*, 19(1):94–102, 2009. (Cited on page 203.)

[469] Josephine H Naish, Lucy E Kershaw, David L Buckley, Alan Jackson, John C Waterton, and Geoffrey JM Parker. Modeling of contrast agent kinetics in the lung using T1-weighted dynamic contrast-enhanced MRI. *Magnetic resonance in medicine*, 61(6):1507–1514, 2009. (Cited on pages 203, 212, and 216.)

[470] E Ashton, E Durkin, E Kwok, and J Evelhoch. Conversion from signal intensity to Gd concentration may be unnecessary for perfusion assessment of tumors using DCE-MRI. In *Proceedings of the 15th Annual Meeting of ISMRM*, page 2813. ISMRM, 2007. (Cited on page 204.)

[471] Simon Buus, Cai Grau, Ole Lajord Munk, Dirk Bender, Kenneth Jensen, and Susanne Keiding. ^{11}C-methionine PET, a novel method for measuring regional salivary gland function after radiotherapy of head and neck cancer. *Radiotherapy and oncology*, 73(3):289–296, 2004. (Cited on pages 212, 215, and 216.)

[472] Adrianus J De Langen, Vivian EM Van Den Boogaart, J Tim Marcus, and Mark Lubberink. Use of $H_2{}^{15}$O-PET and DCE-MRI to measure tumor blood flow. *The oncologist*, 13(6):631–644, 2008. (Cited on pages 212 and 216.)

[473] J Ward, AG Chalmers, AJ Guthrie, M Larvin, and PJ Robinson. T2-weighted and dynamic enhanced MRI in acute pancreatitis: comparison with contrast enhanced CT. *Clinical radiology*, 52(2): 109–114, 1997. (Cited on page 213.)

[474] Anwar R Padhani. Dynamic contrast-enhanced MRI in clinical oncology: Current status and future directions. *Journal of Magnetic Resonance Imaging*, 16(4):407–422, 2002. (Cited on page 213.)

[475] Ping Wang, Yiqun Xue, Xia Zhao, Jiangsheng Yu, Mark Rosen, and Hee Kwon Song. Effects of flip angle uncertainty and noise on the accuracy of DCE-MRI metrics: comparison between standard concentration-based and signal difference methods. *Magnetic resonance imaging*, 33(1):166–173, 2015. (Cited on page 216.)

[476] Michael Haimerl, Niklas Verloh, Claudia Fellner, Florian Zeman, Andreas Teufel, Stefan Fichtner-Feigl, Andreas G Schreyer, Christian Stroszczynski, and Philipp Wiggermann. MRI-based estimation of liver function: Gd-EOB-DTPA-enhanced T1 relaxometry of 3T vs. the MELD score. *Scientific reports*, 4, 2014. (Cited on page 216.)

[477] Sonia C Simon-Zoula, Chris Boesch, Frederik De Keyzer, and Harriet C Thoeny. Functional imaging of the parotid glands using blood oxygenation level dependent (BOLD)-MRI at 1.5T and 3T. *Journal of Magnetic Resonance Imaging*, 27(1):43–48, 2008. (Cited on page 216.)

[478] CR Habermann, MC Cramer, J Graessner, P Gossrau, F Reitmeier, J Fiehler, V Schoder, M Jaehne, and G Adam. Functional imaging of parotid glands: diffusion-weighted echo-planar MRI before and after stimulation. In *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, volume 176, pages 1385–1389. © Georg Thieme Verlag KG Stuttgart· New York, 2004. (Cited on page 216.)

[479] Using Irregular Tessellations. Hierarchical image analysis using irregular tessellations. *IEEE transactions on pattern analysis and machine intelligence*, 13(4):307, 1991. (Cited on page 281.)

[480] NP Weatherill. A method for generating irregular computational grids in multiply connected planar domains. *International Journal for Numerical Methods in Fluids*, 8(2):181–197, 1988. (Cited on page 281.)

Appendix A

The Optimal Obliquity of Cleaving Planes

Oblique cleaving planes help reduce the problem of grid-plane alignment where many voxels (i.e., whole rows, columns, slices) cross the boundary of a sub-segment when the cleaving plane is moved a small amount, as can happen if some contour edges are colinear with the voxel grid. There is an optimal cleave plane orientation that can be derived when sub-segment extents are known exactly. This direction maximizes the minimum spacing between voxel-plane distance, ensuring small changes in the plane position results in the smallest possible number of voxels crossing the plane at one time (e.g., minimizing spatial resonances)[1]. For example, on a planar Cartesian grid considering the origin and nearest-neighbours only the min-max angle for a line[2] (within $[0, \frac{\pi}{4}]$, but seven more angles are identical due to symmetry) is 26.56505°. Including next-nearest neighbours yields 18.43495°. (Both are shown in figure A.1.) The ultimate sub-segment dimensions are not known beforehand, so the angle must be estimated. Unfortunately, even estimation is difficult and costly [435, 436]. Figure A.2) depicts the objective function which must be maximized for square regions of limited extent. To simplify the use of oblique planes in this work, the requirement of optimality was relaxed

[1] Incidentally, this angle also appears to minimize the maximum spacing.
[2] This angle comes from the solution of $2\sin x = \cos x$ within $[0° : 45°]$.

and a cyclic rotation of 22.5° between cardinal axes was used throughout. Results indicate it was appropriate for our data set. It is worth noting that in many cases the problem can be avoided somewhat by irregular tessellation (e.g., [434, 479, 480]) and resampling of the dosimetric grid.

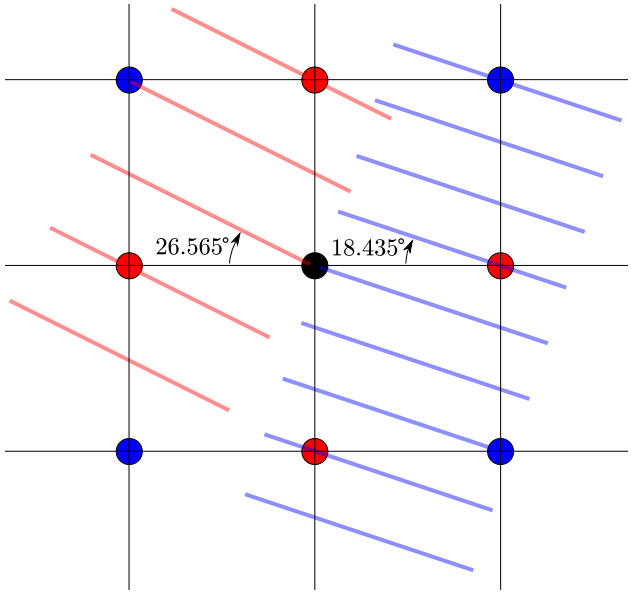


Figure A.1: Depiction of voxel spacing at the min-max angle when nearest-neighbours (origin and red; red lines) and next-nearest-neighbours (origin, red, and blue; blue lines) are included. Figure A.2 depicts how these angles are located.

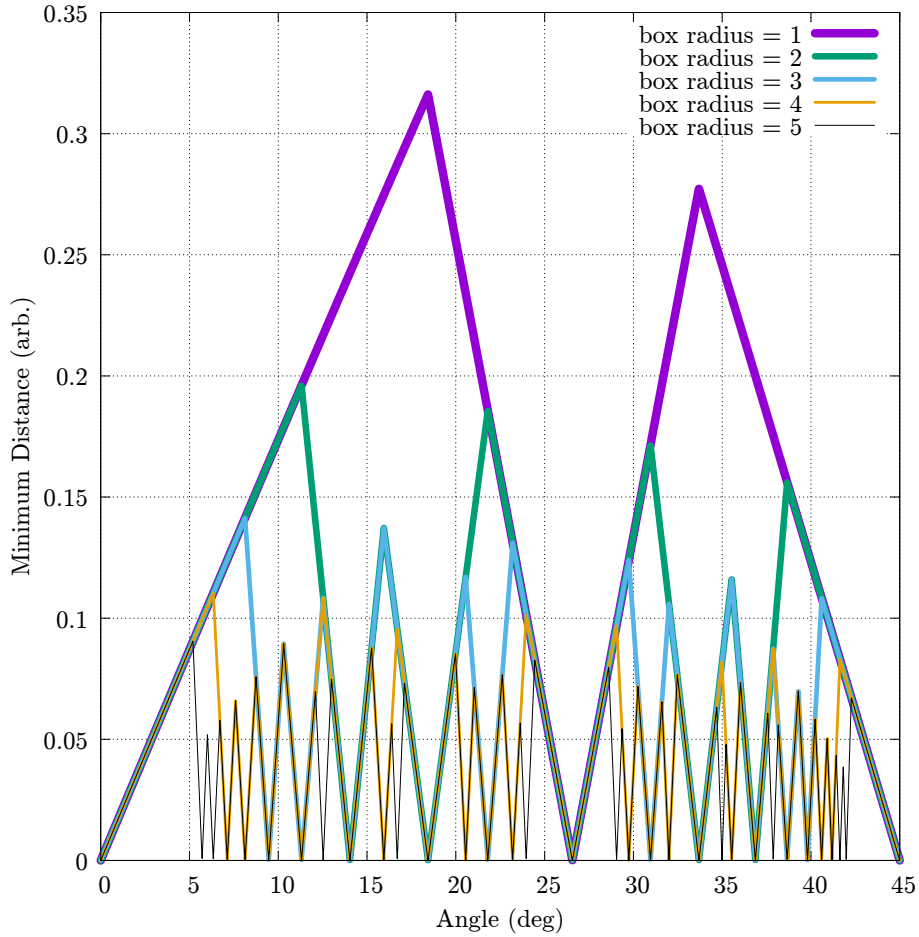


Figure A.2: Minimum spacing between distances of voxel centres above line to line vs line angle for various box radii (i.e., a box of radius n centred at the origin contains $(2n+1)^2$ vertices). The next-nearest-neighbours example of fig. A.1 corresponds to a box radius of 1. Grid spacing is $\Delta x = \Delta y = 1$ (arb. units). The min-max angle is the angle that maximizes this function. For square regions the left-most peak is maximal and shrinks as the box radius grows. For arbitrary geometry (e.g., ROIs) this is no longer generally true.