

DEVELOPMENT OF IMAGING-BASED RESPONSE PREDICTORS FOR PERSONALIZED RADIOTHERAPY IN HEAD AND NECK CANCER



Nienke D. Sijtsema

Development of imaging-based response predictors for personalized radiotherapy in head and neck cancer

Nienke D. Sijtsema

Cover design: Femke van Os

Layout: Jacolijn de Krom | www.persoonlijkproefschrift.nl

Printed by: Ridderprint | www.ridderprint.nl

ISBN: 978-94-6483-967-8

Copyright © 2024 N.D. Sijtsema

All rights reserved. No parts of this thesis may be reproduced or transmitted in any form by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the author, or, when appropriate, from the publishers of the publications.

Development of Imaging-Based Response Predictors for Personalized Radiotherapy in Head and Neck Cancer

Ontwikkeling van op beeldvorming gebaseerde responsvoorspellers voor gepersonaliseerde radiotherapie van hoofd-hals kanker

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof. dr. A.L. Bredenoord

en volgens besluit van het college voor promoties.
De openbare verdediging zal plaatsvinden op

dinsdag 21 mei 2024 om 10.30 uur

door

Nienke Desiree Sijtsema
geboren te Amsterdam.

Erasmus University Rotterdam



Promotiecommissie:

Promotoren:

Prof. dr. M.S. Hoogeman
Prof. dr. J.A. Hernández Tamames

Overige leden:

Prof. dr. R.A. Nout
Prof. dr. U.A. van der Heide
Dr. ir. M.E.P. Philippens

Copromotoren:

Dr. ir. S.F. Petit
Dr. ir. D.H.J. Poot

Voor papa

Voor Ernst

Contents

Chapter 1	Introduction	9
Part one	MRI for tumor response prediction	21
Chapter 2	An optimal acquisition and post-processing pipeline for hybrid IVIM-DKI in head and neck	23
Chapter 3	Multi-delay pseudo-continuous arterial spin labeling to measure blood flow in the head and neck	51
Chapter 4	Correlation between multi-delay pseudo-continuous arterial spin labeling and intravoxel incoherent motion diffusion-weighted imaging in the head and neck	71
Chapter 5	The COMPLETE trial: HolistiC early respOnse assessMent for oroPharyngeal cancer paTiEnts; Protocol for an observational study	93
Chapter 6	Relating pretreatment non-Gaussian intravoxel incoherent motion diffusion-weighted imaging to HPV status and response in oropharyngeal carcinoma	109
Part two	Response prediction of organs at risk	129
Chapter 7	Accounting for fractionation and heterogeneous dose distributions in the modeling of osteoradionecrosis in oropharyngeal carcinoma treatment	131
Chapter 8	Development of a local dose-response relationship for osteoradionecrosis within the mandible	151
Chapter 9	Discussion	169
	Summary	182
	Samenvatting	187
	References	193
	List of publications	213
	PhD portfolio	215
	Curriculum Vitae	217
	Dankwoord	219

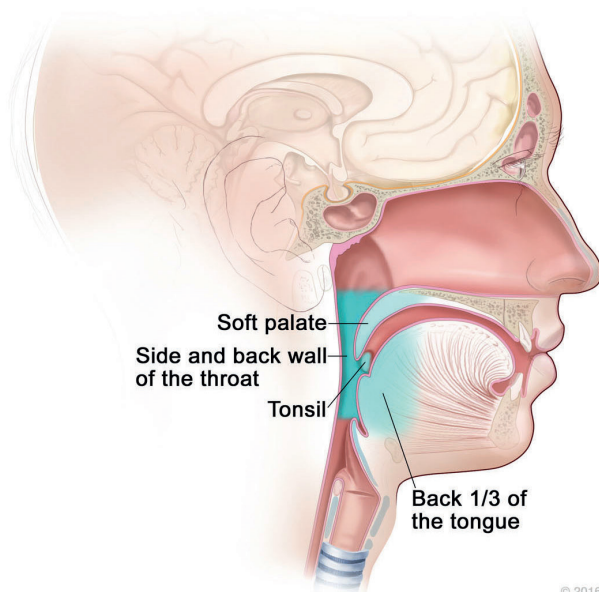
1

Introduction

1.1 Head and neck cancer, treatment, and risk factors

Head and neck cancer represents around 5% of the new cancer incidences worldwide [1], roughly translating to a million new cases per year. In the Netherlands, annually around 20% of the head and neck cancers arise from the oropharynx, and the incidence has been rising slowly in the last few decades [2], with common subsites being the tonsils, base of tongue, oropharyngeal wall, and soft palate (Figure 1.1). Common risk factors for oropharyngeal squamous cell carcinoma (OPSCC) include smoking and alcohol use. In fact, heavy usage of both tobacco and alcohol can increase the risk of OPSCC by over 35 times [3]. Additionally, human papillomavirus (HPV) infection is an important risk factor, and the prevalence of HPV-related cancers within the group of OPSCC has been rising rapidly over the last few decades from roughly 40.5% before 2000, to 72.2% between 2005 and 2009 [4]. Especially in recent years, it has been recognized that within the group of OPSCC, we should distinguish two separate diseases, namely HPV-related and non-HPV-related OPSCC [5, 6]. The major reason for this distinction is the clearly different prognostic pattern [7-9] and distinctive histopathological features [10-12]. To reflect this difference in outlook for HPV-related and non-HPV-related OPSCC, new, separate staging systems have emerged for these two groups during the last 5 years [13].

Parts of the Oropharynx



© 2016 Terese Winslow LLC

Figure 1.1. The anatomical location of the oropharynx, with common cancer subsites; the tonsil, base of tongue, oropharyngeal wall, and soft palate. For the National Cancer Institute ©2016 Terese Winslow LLC, U.S. Govt. has certain rights, used with permission.

Conventional treatment of OPSCC includes surgery, chemotherapy, radiotherapy, or a combination thereof, depending on the location, stage, and patient vitality. Patients receiving (chemo)radiotherapy as the primary mode of treatment are commonly treated with 35 fractions of 2 Gy (70 Gy in total) to the primary tumor and pathological lymph nodes in the neck. The neck can also be irradiated electively, in which case a dose of 54.25 Gy is commonly prescribed to this area. Generally, 5 to 6 fractions per week are delivered, so a radiotherapy course of 35 fractions can take up to 7 weeks to complete. In light of the COVID-19 pandemic, there is a renewed interest in hypofractionated treatments, reducing the amount of fractions, and in most cases, overall treatment time [14-16].

The head and neck is a challenging area for the application of radiotherapy due to the presence of many important healthy tissues, called organs at risk (OARs). Unfortunately, these can usually not be fully spared during treatment and may receive significant dose, which in turn could give rise to side effects (toxicity). Xerostomia (dry mouth) due to irradiation of the salivary glands is a commonly observed toxicity in OPSCC [17]. Additionally, dysphagia (difficulty swallowing) [18] due to irradiation of the swallowing muscles as well as loss of taste due to irradiation of gustatory organs [19] are commonly observed side effects that are also adversely affected by the presence of xerostomia [20, 21]. These are usually side effects that develop during or within 3 months after radiotherapy, also called acute toxicity. While acute toxicity tends to diminish after treatment, there is a substantial group in which toxicity becomes chronic. Yet, some side effects develop months or years after radiotherapy [22]. One important severe late side effect is osteoradionecrosis of the jaw, where bone tissue is dying due to the effects of radiation [23]. It can cause pain, oral skin fistulae, and even fractures requiring extensive reconstruction. In general, toxicities can have a significant negative impact on the quality of life of the patient [18, 24].

While delivering a sufficient dose to the primary tumor and affected lymph nodes is usually the primary objective, the risk of toxicity should also be acceptable. Therefore, dose constraints are applied to key healthy tissues to avoid unnecessary dose [25]. In practice, tumor dosage is sometimes slightly compromised to keep the risk of toxicity within an acceptable range. Generally, the same dose to tumor and dose constraints to healthy tissue are deployed for all patients within a certain treatment group. However, tumor and healthy tissue response to radiation differ per patient, and therefore not all tumors are cured and the severity of toxicity varies substantially per patient. Insight into patient-specific radiosensitivity could be used to adjust treatment and in turn, improve cure rate and/or reduce toxicity in individual patients. To achieve

this, approaches are required that predict treatment response of both tumor and healthy tissues in individual patients. Therefore, this thesis focuses on developing prediction markers and prediction models for personalization of radiation therapy. Figure 1.2 schematically depicts how each chapter relates to the prediction of tumor response and tolerance of OARs. In section 1.2 the MRI techniques used in this thesis are explained. In section 1.3 dose-volume-based prediction modeling for organs at risk is discussed.

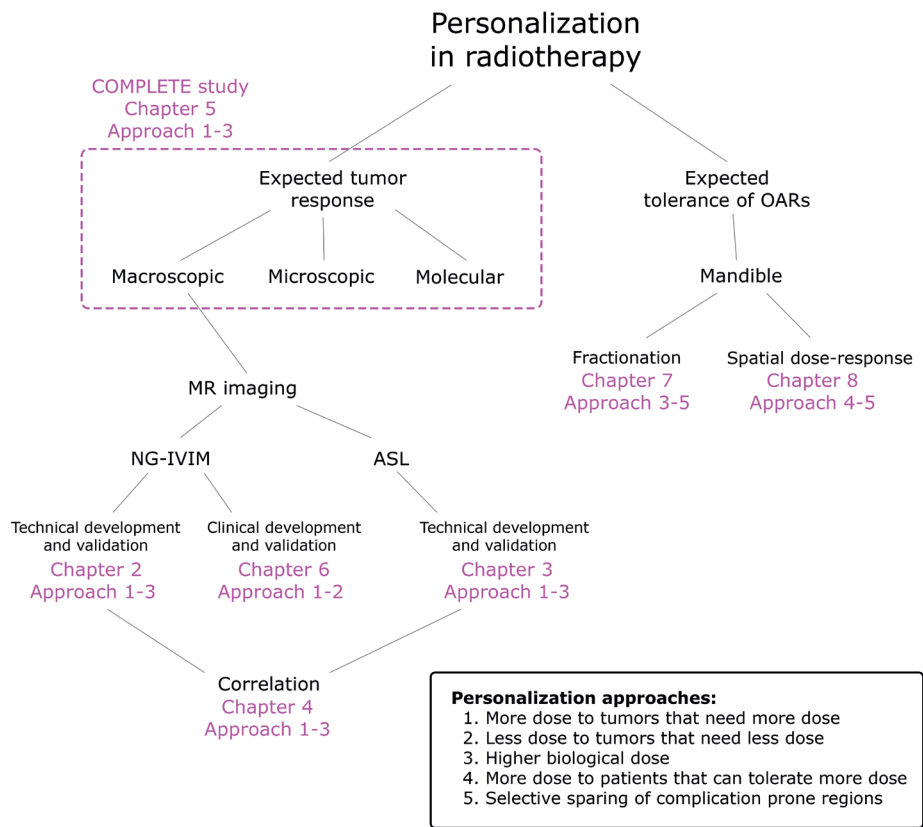


Figure 1.2. Personalization in radiotherapy depends on the expected tumor response on a macroscopic, microscopic, and molecular level (dashed box), as well as the expected tolerance of OARs. For the prediction of tumor response, this thesis focuses on implementing and optimizing non-Gaussian intravoxel incoherent motion imaging (NG-IVIM) and arterial spin labeling (ASL) for the prediction of the macroscopic tumor response. Additionally, the correlation between these two techniques was investigated. For the prediction of the tolerance of OARs, we specifically focus on the mandible and investigated the influence of fractionation and spatially varying dose on the risk of osteoradionecrosis.

1.2 Imaging for tumor response prediction and monitoring

There are several promising ways to predict and monitor response to treatment, both for healthy tissue and tumors alike. One possibility is to use biological quantitative magnetic resonance imaging (MRI) techniques. While anatomical images tell us about the location of tissues and tumors, with biological MRI we can measure different tissue properties, like diffusion, cellularity, and perfusion, that are typically disrupted in cancers compared to healthy tissues as is outlined in detail in papers by Hanahan and Weinberg [26–28] on the hallmarks of cancer. Biological quantitative MRI techniques therefore could allow us to gain insight into several hallmarks of cancer, such as disrupted vasculature or tumor-promoting inflammation.

In this thesis, we consider diffusion-weighted imaging (DWI) and arterial spin labeling (ASL) as the most promising techniques for response assessment in the head and neck, as explained in the following sections. However, many more MRI techniques exist that could give us information about the tissue response to treatment in the head and neck, such as dynamic contrast enhanced (DCE) MRI [29, 30], Chemical Exchange Saturation Transfer (CEST) MRI [31–33] and MRI spectroscopy [34].

1.2.1 Diffusion-weighted imaging

Introduced in 1965 by Stejskal and Tanner [35], DWI is a versatile technique that yields parameters related to diffusion, tissue density, cellularity, and even perfusion properties of tissue on the micro-environment level, depending on the settings and model used. Tightly packed, dense tissues, with many cells, will show less diffusion than tissues with larger cells and more extracellular space. While tumors tend to have high tissue density, if the treatment successfully eradicates tumor cells, the tissue will become less dense and the diffusion in the tissue will likely increase. So, an increase in diffusion parameters measured with DWI could indicate the tumor is responding to the treatment. Therefore, DWI is of high interest for response prediction, and several groups have confirmed its potential as a response predictor [36–38]. Since DWI is sensitive to all moving water in the micro-environment, it can also yield measures related to microvascular perfusion, which is of interest because an increase in perfusion is associated with angiogenesis, one of the hallmarks of a growing tumor. So, increasing perfusion during treatment could indicate a non-responding tumor. On the other hand, decreasing perfusion during treatment could indicate the destruction of cells, leading to a reduced demand for oxygen and nutrients, indicating a responding tumor.

The key setting in DWI is the b-value, which controls the size (including strength and length) of the diffusion weighting gradients and therefore determines the amount of diffusion weighting in the image. Typically, diffusion images also consider the diffusion direction, although this is mostly relevant in tissues with high directionality like brain and muscle tracts, but is not deemed relevant in the case of non-directional tissues like tumors or the salivary glands.

At low b-values (<200 s/mm²), signal is lost only from fast-moving spins (i.e., those related to perfusion), while at high b-values (>1000 s/mm²) only the signal from very slow-moving spins is left (i.e., those related to intracellular-diffusion). For the head and neck, we roughly distinguish three regimes depending on b-values: (1) from 0-200 s/mm²: perfusion, (2) from 200-1000 s/mm²: extra-cellular diffusion, (3) >1000 s/mm²: intracellular diffusion [39]. In its simplest form only extra-cellular diffusion is assessed through estimation of the apparent diffusion coefficient (ADC) in the following mono-exponential model:

$$S_i = S_0 e^{-b_i ADC} \quad (1.1)$$

where S_i is the measured signal intensity at the corresponding b-value b_i and S_0 the signal intensity at a b-value of 0 s/mm². However, ADC also contains the fast (perfusion) component of diffusion, and can therefore be biased. To eliminate this bias, the model can be extended to the bi-exponential intravoxel incoherent motion (IVIM) model, introduced by Le Bihan [40, 41] in 1986. In the IVIM model, perfusion is assessed by estimating the perfusion fraction f , as well as the pseudo-diffusion coefficient D^* . Additionally, the diffusion coefficient D (analogous to the ADC) is estimated by the following model:

$$S_i = S_0 \left((1 - f) e^{-b_i D} + f e^{-b_i D^*} \right) \quad (1.2)$$

To quantify the restrictedness of the diffusion, the IVIM model can be even further extended [42] by adding Kurtosis K to assess the intracellular diffusion to the Non-Gaussian IVIM (NG-IVIM) model, also called IVIM-DKI (intravoxel incoherent motion-diffusion kurtosis imaging):

$$S_i = S_0 \left((1 - f) e^{-b_i D + \frac{1}{6} (b_i D)^2 K} + f e^{-b_i D^*} \right) \quad (1.3)$$

Generally, the amount of b-values and the distribution over the regimes depends on the preferable model. For example, only 2-3 b-values would be sufficient to estimate the ADC, while many more are needed for the IVIM and NG-IVIM models. A full image is acquired at each chosen b-value. Therefore, all estimates can be done on a voxel-wise basis, so it yields maps of the

estimated parameters. As a consequence, DWI has the ability to also assess tissue heterogeneity.

However, DWI can be challenging since it commonly uses echo planar imaging (EPI) due to the fast read-out required for imaging in the water motion regime. EPI is particularly sensitive to inhomogeneity in the static magnetic field (B_0) which is inherently introduced due to susceptibility differences in the imaged subject. Specifically, inhomogeneity in the B_0 field can introduce geometric deformation of millimeters up to centimeters in the image. Yet, in radiotherapy, a highly accurate geometry is required, especially in this era where highly conformal radiotherapy techniques are commonly used. So, EPI in the head and neck, an area with strong differences in susceptibility can lead to unacceptable distortion in the radiotherapy context [43]. Although efforts have been made to develop pulse sequences that are more robust to geometrical distortion [44–46], they are not widely available at this moment as availability often depends on the vendor. Another option is to correct for geometrical distortion. Since the direction of the distortions in EPI depends on the phase encoding direction, using an extra acquisition with opposite phase encoding direction can be used to estimate a B_0 field map and correct geometrical distortion [47, 48].

Aside from this, DWI is prone to motion artifacts, since it is sensitive to the small motion of water molecules [49]. Head motion in radiotherapy is usually limited using an immobilization mask, but nevertheless, small movements along a substantial scan time can lead to slightly different head positions in each b-value, thereby impacting quantification. Additionally, other motions specific to the head and neck area, such as swallowing, coughing, or even sneezing, can occur. A relatively complex, and quick motion like swallowing usually presents as signal dropout and could therefore significantly impact image quality and quantification.

1.2.2 Arterial spin labeling

ASL is an MRI technique that measures perfusion, more specifically the blood flow to a certain tissue [50]. As mentioned in the section on DWI, increasing perfusion could be a sign of angiogenesis, a well-known characteristic in growing tumors. On the other hand, decreasing perfusion could indicate the tumor is responding to treatment. Studies on ASL in the head and neck focusing on response prediction are currently scarce, but Fujima et al. demonstrated the potential of ASL as a response predictor [51]. While DWI is also capable of measuring perfusion-related characteristics, it is currently unclear whether ASL and DWI describe the same aspects of tissue perfusion, and it is currently unknown which aspect of tissue perfusion will best reflect response.

In contrast to other MRI techniques that measure perfusion, ASL does not require injection of a contrast agent. Instead of using an external contrast agent, the blood is magnetically tagged. To this end, a labeling plane is placed in the neck. In this plane, all protons are labeled by an inversion radiofrequency pulse. Subsequently, the blood flows to the tissue of interest, which takes a certain time also known as the arterial transit time, before the tissue is imaged. The presence of labeled blood in a tissue causes a signal decrease compared to a non-labeled image.

Subtracting the labeled image from the control (non-labeled image) yields a perfusion-weighted image (Figure 1.3) from which the blood flow can be quantified. Like with DWI parameters, blood flow can be fitted or calculated voxel-wise and a map of the blood flow is obtained that allows for assessment of tissue heterogeneity. In conventional ASL, typically only one post-label delay (the time between labeling and imaging) is used. Ideally, the post-labeling time is the same as the arterial transit time, and the maximum signal decrease is achieved exactly at the time of imaging. In practice, however, the arterial transit time, and therefore the ideal post-labeling delay, is often unknown. It varies per tissue, as well as per individual. In this thesis, we therefore use multi-delay encoded ASL. This novel technique allows for imaging at multiple post-label delays, resulting in a blood flow and arterial transit time estimate. As a result, the blood flow estimate is not biased by arterial transit time. Additionally, it allows for simultaneous assessment of tissues that have inherently different arterial transit times, for example due to different vascular path lengths.

1.3 Modeling for risk of toxicity prediction

Several factors could affect a patient's risk of developing side effects. Partly, they could be similar to prognostic factors for tumor response, such as smoking and alcohol use, but they might also be specific to the OAR, such as dose to the OAR, or the functionality of the OAR at baseline or early during treatment. For common toxicities such as dysphagia and xerostomia, well-established prediction models exist, such as those used in the selection of head and neck cancer patients for proton therapy in the Netherlands [53]. However, for osteoradionecrosis of the lower jaw, which is a very severe toxicity though relatively rare, prediction models are currently lacking due to a couple of reasons.

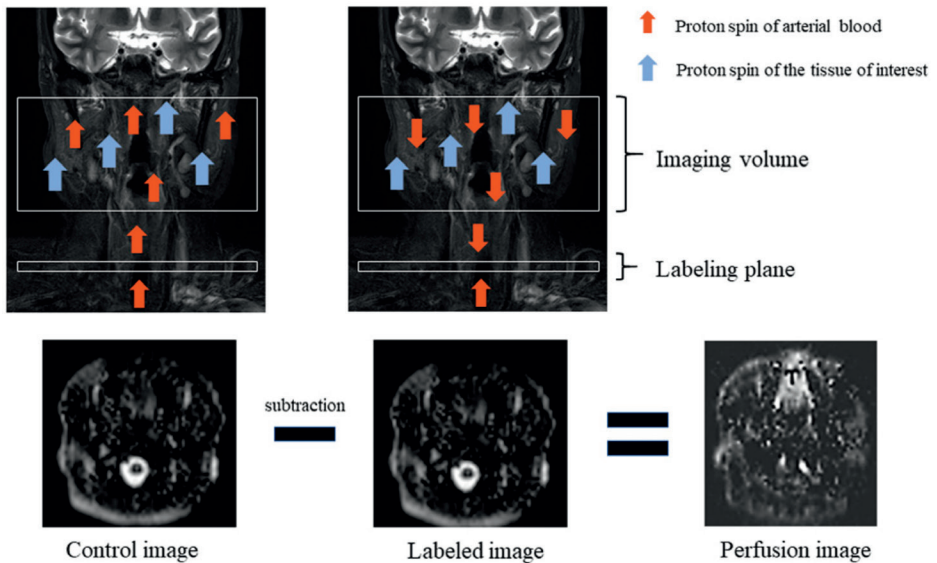


Figure 1.3. Schematic of the acquisition of arterial spin labeling in the head and neck taken from Tanaka et al. [52] under CC BY license. A labeling plane is placed in the neck. In this plane, all protons are labeled by an inversion radiofrequency pulse. Subsequently, the blood flows to the tissue of interest. The presence of labeled blood in a tissue causes a signal decrease compared to a non-labeled image. Subtracting the labeled image from the control (non-labeled image) yields a perfusion-weighted image.

First, determining the relevant dose parameter to include in the prediction model is not straightforward. The radiation dose received by the mandible is heterogeneous and many different dose parameters can be chosen such as dose-volume histogram parameters, mean and maximum dose, and so on. Each of these reflects some aspect of the 3D dose distribution, but the spatial information is not retained. Secondly, the mandible can experience local trauma, for example in the form of extractions which were repeatedly linked to an increased risk of osteoradionecrosis [54, 55]. This raises the question of whether local modeling of toxicity risk could help to identify and steer dose away from more radiosensitive areas. However, developing local dose-response relationships requires complex modeling procedures and the identification of relevant subvolumes in each patient. Lastly, ORN is a rare toxicity, and large groups of patients are needed to produce an externally validated model. In large groups, patients are often irradiated with a variety of treatment techniques which cannot always be compared directly in terms of dose-volume histogram parameters, because the dose distribution characteristics can differ significantly. This is the case for example when

comparing conventional radiation schemes with hypofractionated stereotactic boost schemes.

1.4 Aim and outline of this thesis

The primary objective of this thesis is to develop imaging-based markers that provide insight into the prediction of tumor response and treatment-related toxicity. By using these markers in the future, we aim to optimize treatment planning and decision-making, ultimately leading to improved patient outcomes. Within the scope of this thesis, DWI and ASL are investigated as promising candidates for response prediction biomarkers, and the prediction of treatment-related toxicity is studied in the context of osteoradionecrosis.

The first part of this thesis focuses on using MRI for tumor response prediction. In **Chapter 2** we optimize the acquisition parameters of DWI for the NG-IVIM model specifically for the head and neck region. We evaluate it in a cohort of healthy volunteers and compare it to acquisition parameters from literature. In that chapter, we also evaluate the influence of intra- and inter-b-value registration of the DWI images and the presence of signal dropout due to swallowing on the DWI quantification. In **Chapter 3** we evaluate multi-delay pseudo-continuous ASL in the parotid glands, submandibular glands, tonsils, and cerebellum of healthy volunteers with the aim to explore nominal values of blood flow. Additionally, we look into the influence of T1 and transit-time correction on the quantification of the blood flow to enable accurate blood flow measurements of multiple tissues simultaneously. Since both ASL and (NG-)IVIM provide measures of perfusion of tissue, if they provide the same information it would not be necessary to acquire both scans, therefore, in **Chapter 4**, we explore the correlation between parameters obtained from multi-delay pCASL and IVIM. The optimized NG-IVIM protocol obtained from **Chapter 2** is implemented in a clinical study investigating the ability of NG-IVIM (among others) as a predictor for tumor response. **Chapter 5** outlines the protocol of this clinical study, which, aside from assessing macro-scoping response with NG-IVIM, also acknowledges the potential of microscopic and molecular response assessment. **Chapter 6** concludes the first part of this thesis and explores the application of the NG-IVIM sequence presented in **Chapter 2** in oropharyngeal cancer patients with a focus on the effect of HPV status on NG-IVIM parameter values.

The second part of this thesis focuses on response prediction in OARs. Specifically, in **Chapter 7**, we investigate the risk factors and dose response for osteoradionecrosis in the lower jaw on a patient level, as well as which dose parameters are suitable to include in a risk model in case of heterogeneous

radiotherapy treatment schedules. In **Chapter 8**, we build a local dose-response model for osteoradionecrosis of the mandible. This model takes into account patient-level characteristics, local dose, the location of pre-radiotherapy extractions (trauma), and the location of osteoradionecrosis while also taking into account the non-independence of dental elements of the same patient.

Finally, in **Chapter 9**, the work presented in this thesis is discussed in a wider context, including remaining challenges and future perspectives.

Part One

2

An optimal acquisition and post-processing pipeline for hybrid IVIM-DKI in head and neck

Nienke D. Sijtsema, Steven F. Petit, Dirk H.J. Poot,
Gerda M. Verduijn, Aad van der Lugt,
Mischa S. Hoogeman,
Juan A. Hernandez-Tamames

This chapter is based on Sijtsema ND, Petit SF, Poot DHJ, Verduijn GM, van der Lugt A, Hoogeman MS, et al. An optimal acquisition and post-processing pipeline for hybrid IVIM-DKI in head and neck. Magn Reson Med. 2021;85:777-89.

Abstract

Purpose

To optimize the diffusion-weighting b-values and post-processing pipeline for hybrid intravoxel incoherent motion diffusion kurtosis imaging (IVIM-DKI) in the head and neck region.

Methods

Optimized diffusion-weighting b-value sets, ranging between 5 and 30 b-values, were constructed by optimizing the Cramér-Rao lower bound of the hybrid IVIM-DKI model. With this model f , the perfusion fraction, D^* , the pseudo-diffusion coefficient, D , the diffusion coefficient, and K , the kurtosis, were estimated. Sixteen volunteers were scanned with a reference b-value set and three repeats of the optimized sets, of which one with volunteers swallowing on purpose.

The effect of (1) b-value optimization and number of b-values; (2) registration type: none versus inter-volume versus intra- and inter-volume registration and (3) manual swallowing artifact rejection on the parameter precision was assessed.

Results

The standard deviation (SD) was higher in the reference set for f , D , and K by a factor of 1.7, 1.5, and 2.3 compared to the optimized set, respectively. A smaller SD (factor 0.7) was seen in D^* . The sets containing 15, 20, and 30 b-values had comparable repeatability in all parameters, except D^* for which set size 30 was worse. Equal repeatability for the registration approaches was seen in all parameters of interest. Swallowing artifact rejection removed the bias when present.

Conclusion

To achieve optimal hybrid IVIM-DKI imaging in the head and neck region, b-value optimization and swallowing artifact image rejection are beneficial. The optimized set of 15 b-values yielded the optimal protocol efficiency, with a precision comparable to larger b-value sets and a 50% reduction in scan time.

2.1 Introduction

In the last decade, the apparent diffusion coefficient has been shown to be a promising parameter for response assessment of head and neck cancer treated with (chemo)radiotherapy [36, 56]. More recently, the benefit of obtaining additional parameters from diffusion-weighted imaging (DWI) by intravoxel incoherent motion (IVIM) and diffusion kurtosis imaging (DKI) has been acknowledged [42, 57]. Combining these models in hybrid IVIM-DKI allows for the simultaneous assessment of the diffusion coefficient (D), perfusion fraction (f), pseudo-diffusion coefficient (D^*) and kurtosis (K). Both f and D^* are related to perfusion and could therefore provide insight into abnormal tumor perfusion. The latter parameter, K , is associated with restricted diffusion and cellularity [58], which is commonly increased in tumors. However, full assessment of their potential as response biomarkers and, in the future, for application in radiotherapy treatment planning and treatment adaptation, requires robust and repeatable estimation of all IVIM-DKI parameters within a patient [59].

IVIM-DKI parameter estimation is currently hampered by inefficient sampling of the DWI signal curve and motion artifacts. Inefficient sampling of the DWI signal curve, due to non-optimal choice of b-values, leads to unnecessary long scan times in order to achieve a similar precision as optimized sampling. Moreover, long scan times might increase the amount of motion corruption. Several efforts have been made to optimize b-values for both the mono-exponential [60–66], the IVIM model with direct [66–69] and segmented fitting [70], and the kurtosis model [66] in a variety of tissues, although none of the mentioned papers address the head and neck region specifically. For complex models, such as hybrid IVIM-DKI, optimizing b-values is increasingly more difficult and, to the best of our knowledge, has not been done so far.

Additionally, the head and neck region is prone to several types of motion artifacts. Firstly, swallowing and coughing artifacts, which present themselves as severe signal dropout, could cause over- or underestimation of the DWI parameters. Current mitigation strategies mainly consist of specific patient instruction for not swallowing or coughing, but are not always sufficient because both can be a reflex behavior. Secondly, head motion hampers parameter estimation by causing misalignment between scan volumes. This type of motion can be partially prevented by using fixation of the patient; either in the form of padding, or, in the case of radiotherapy planning MRI, with an immobilization mask. Additionally, motion artifacts might be corrected after acquisition by registration; however, registration of high b-value images can be problematic due to lower signal-to-noise ratio (SNR).

Therefore, the goal of this study was to find an optimal acquisition and develop a post-processing pipeline for hybrid IVIM-DKI DWI in the head and neck region. To this end, we optimized the b-values of hybrid IVIM-DKI for the head and neck region, applied motion compensation, and investigated the effect of swallowing artifact rejection.

2.2 Methods

Acquisition optimization and post-processing of the hybrid IVIM-DKI DWI data consisted of three stages. In the first stage, described in section 2.2.1, the set of b-values was optimized for different b-value set sizes based on simulated ground truth voxels. Next, these b-value sets were scanned in healthy volunteers, as described in section 2.2.2. In the second stage, intra- and inter-volume registration were applied to the acquired data as described in section 2.2.3. The third stage consisted of swallowing artifact image rejection (section 2.2.4). Lastly, section 2.2.5 describes the assessment of the parameter estimation precision of the pipeline. In this paper, the term post-processing refers to all processing done after acquisition, but before parameter estimation.

2.2.1 Optimization of b-values

The hybrid IVIM-DKI model is given by:

$$S_b(\boldsymbol{\theta}) = S_0 \left((1 - f) e^{-bD + \frac{1}{6}(bD)^2 K} + f e^{-bD^*} \right) \quad (2.1)$$

where $S_b(\boldsymbol{\theta})$ is the signal at a specific b-value (amount of diffusion weighting) as a function of the parameters $\boldsymbol{\theta} \in \{S_0, f, D^*, D, K\}$ which are described in Table 2.1.

Table 2.1. Overview of parameters in the hybrid IVIM-DKI model and the minimum and maximum values chosen for the ground truth value ranges. S_0 in this table is the estimated signal intensity without T2 decay effects.

Parameter	Description	Minimum	Maximum
S_0	Signal intensity at $b=0$ s/mm ²	4275.8	7126.3
D	Diffusion coefficient	$0.25 \cdot 10^{-3}$ mm ² /s	$3.41 \cdot 10^{-3}$ mm ² /s
f	Perfusion fraction	0.09	0.42
D^*	Pseudo-diffusion coefficient/ Apparent perfusion coefficient	$6.29 \cdot 10^{-3}$ mm ² /s	$237.39 \cdot 10^{-3}$ mm ² /s
K	Kurtosis	0.1	2.81

A ground-truth set $\boldsymbol{\theta}$ was created using a Halton sequence [71] for the ground-truth ranges shown in Table 2.1. A Halton sequence yields a sequence of low-discrepancy, pseudo-random numbers. These ranges were derived from the parameter values reported in several previous studies in head and neck tumors [42, 56, 72-74]. From these ground-truth parameter sets, sets showing a D^* smaller than 6 times D were removed to be able to properly separate the two parameters. Some parameter combinations result in increased signal at high b-values. To eliminate this nonphysical result, parameter sets in which the partial derivative of the model to the b-value was larger than 0 at $b=1500$ s/mm² were removed. After exclusion, 272 of 576 ground truth $\boldsymbol{\theta}$ were left.

The IVIM-DKI b-value set was optimized for the precision of parameter estimation within the available acquisition time with the cost function:

$$C(\mathbf{b}) = \sqrt{(\text{relCRLB}(\boldsymbol{\theta}, \mathbf{b}))^2}_{\boldsymbol{\theta}, i} \cdot T_{acq}(\mathbf{b}) \quad (2.2)$$

where C is the cost function value, \mathbf{b} is the set of b-values, and the bar with subscripts indicates the mean over parameters i and $\boldsymbol{\theta}$, where i corresponds to f , D^* , D , and K .

In Equation 2.2 the relative Cramér-Rao Lower bound (relCRLB) is

$$\text{relCRLB}(\boldsymbol{\theta}, \mathbf{b}) = (\mathbf{F}(\boldsymbol{\theta})^{-1})_{i,i} / \boldsymbol{\theta}_i^2 \quad (2.3)$$

with $\mathbf{F}(\boldsymbol{\theta})$, the Fisher's Information matrix:

$$F_{i,j}(\boldsymbol{\theta}) = -\mathbb{E}_{\mathbf{S}} \left[\frac{\partial^2 \ln p(\mathbf{S}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right] \quad (2.4)$$

where and $p(\mathbf{S}|\boldsymbol{\theta})$ is the joint probability density function of the Rician distributed measurements \mathbf{S} .

The CRLB, $\mathbf{F}(\boldsymbol{\theta})^{-1}$, gives a lower bound of the variance for the given parameter vector and is commonly used in experiment design [75]. Therefore, it is used as a measure of precision in this paper. The optimization of $C(\mathbf{b})$ was constrained to avoid negative b-values and b-values higher than $b=1500$ s/mm². A maximum b-value of 1500 s/mm² was chosen due to limited SNR in the head and neck region at higher b-values in healthy tissue at 1.5T. The acquisition time was given by $T_{acq} = (TE + RO) \cdot n_{b-values} \cdot n_{slices}$ with $TE = 61 + 0.015 \cdot b_{max}$ and RO the read-out time. This relation between the maximum b-value and TE was

experimentally obtained from the sequence implementation. The read-out time RO was determined to be 120 ms for the scan protocol used in this paper.

Furthermore, a TE correction was applied to the signal S_0 of the ground truth voxels as the signal decreases when TE increases:

$$S_{0,corrected} = S_0 \cdot e^{-\frac{TE}{T_2}} \quad (2.5)$$

The T_2 was set to 80 ms. A maximum b-value of 1500 s/mm² resulted in a modeled signal intensity at b=0 s/mm² of 1500 to 2500. The noise level was set to 20, leading to an estimated intra-tumor SNR of 75–125 at b=0 s/mm². Initial sets were chosen to contain b-values along the range of 0–1500 s/mm². These sets were optimized using the `fminsearch` algorithm in MatLab 2017a (MathWorks, Natick, MA, USA). Optimization was carried out for a set size of 5, 6, 10, 15, 20 and 30 b-values. The set of 20 b-values was optimized first and initialized with approximately linearly distributed b-values. Because the final b-values were clustered, we subsequently restarted the optimization with varying initial distributions of b-values over the three regimes (perfusion (b~0–200 s/mm²), free diffusion (b~200–800 s/mm²), and restricted diffusion b>~800 s/mm²) to reduce the chance of ending in a local minimum. For each number of b-values, the set with the lowest overall cost value was selected. The b-value optimization code is available online at github.com/nsijtsema/IVIMDKI_b-value_optimization.

2.2.2 MR scanning

Seventeen healthy volunteers (14 females, 3 males, mean age 26, age range 19–59) were imaged on a 1.5 T GE Optima MR450w GEM imaging system (General Electric Medical Systems, Waukesha, WI) with a 16-channel head and neck coil (General Electric Medical Systems, Waukesha, WI). The study was approved by the institutional review board (protocol 2014–096) and written informed consent was obtained from all volunteers. The imaging protocol consisted of a T2 PROPELLER (scan time 5 min 34 s) followed by a single shot SE-EPI IVIM-DKI DWI acquisition (FOV 26x26 cm², 128x128 matrix, 2x2x4 mm voxels, 0.2 mm inter-slice gap, TE=81.8 ms, TR=6700 ms, SENSE parallel imaging acceleration factor 2, number of averages = 1, 3 orthogonal diffusion directions) with a reference b-value set (scan time 6 min 35 s) of the geometrical form (0, 10 α^1 , ..., 10 α^{n-1}) with $\alpha=1.3$ and the number of b-values $n=20$, similar to the approach used by Lu et al. [42]. Next, three repeats of the optimized single shot SE-EPI IVIM-DKI DWI sets (scan time 11 min 17 s) were scanned with the same imaging parameters mentioned for the reference b-value set, with exception of the b-values. Volunteers were asked to lie still, except for the last optimized IVIM-DKI acquisition. During that acquisition, volunteers were asked to swallow

four times, so that most optimized b-value sets had at least one swallowing artifact. For all DWI acquisitions, reversed readout gradient polarity for $b=0$ s/mm² (scan time 1 min 7 s) was collected for the purpose of distortion correction using reversed gradient polarity blips [47, 76], which were applied to all images. The odd and even slices of the DWI were acquired in an interleaved fashion to minimize slice crosstalk. Total scan time of the protocol was 55 minutes.

To fit the acquisition in the available scan time, the optimized b-value sets containing 5, 6, 10, 15, 20, and 30 b-values were combined into one set. This was done by adding the b-values of the longer subsequent set that were not already present in the shorter set to the end of the acquisition. The set of 5 was thus expanded with 2 b-values not initially present in that set, to be able to form the optimized set of 6. The set was then consecutively expanded to 10, 15, 20, and 30 b-values in the same manner. Consequently, the b-value sets were grouped in time to ensure the effect of motion is representative of a patient scan. Due to the overlap between the b-values in the optimized sets, the total b-value set encompassed 34 b-values. Table 2.2 contains the acquired b-value sets.

Table 2.2. Overview of the order in which the b-values were acquired for the reference acquisition and the combined optimized IVIM-DKI acquisition.

	b-values [s/mm²]
Reference	0, 10, 1460, 20, 1120, 20, 870, 30, 670, 40, 510, 50, 390, 60, 300, 80, 230, 110, 180, 140
Optimized	0, 20, 780, 1500, 130, 790, 640, 80, 1500, 570, 770, 770, 80, 1500, 780, 1500, 10, 790, 1500, 790, 1500, 80, 750, 1500, 80, 760, 790, 80, 750, 280, 1500, 80, 790, 10

2.2.3 Registration

Two registration methods were compared with non-registering: inter-volume registration only and intra- and inter-volume registration. The methods for both intra- and inter-volume registration were obtained from Guyader et al. [77]. In case of intra-volume registration, the odd and even slices from each b-value image were separated into two separate volumes. Subsequently, the two volumes were aligned by group-wise registration because the odd and even slices were acquired interleaved. Inter-volume registration was carried out by pair-wise registration of each b-value image to the $b=0$ s/mm² image. Intra-volume registration, if applicable, was performed before inter-volume registration. All registrations were non-rigid and carried out with elastix (version 4.9) [78]. The parameter files from the registration approach [77] are available in the elastix parameter file database. Registration was applied

after distortion correction for all cases in the mentioned order. Inter-volume registration errors were detected by manually assessing the imaging volumes and identifying the volumes with severe anatomical mismatches. To correct for registration errors, the transformations of the two well-registered b-values that were closest in time were linearly interpolated and applied to the original image that contained the registration error.

Finally, the $b=0$ s/mm² scan of each acquisition was registered to the reference $b=0$ s/mm² acquisition using the same registration approach as used for the inter-volume registration. The transformation of the $b=0$ s/mm² was propagated to the remaining images of the acquisition.

2.2.4 Swallowing artifact image rejection

Swallowing artifacts presented as severe signal dropout in the individual b-value images were detected by visual inspection and subsequently rejected. Because the data were acquired using 3 sequential orthogonal gradient directions, artifacts present in one diffusion direction resulted in rejection of all three directions to maintain isotropic weighting in the fit.

2.2.5 Assessment

Regions of interest (ROIs) were drawn inside both tonsils of each volunteer based on the first acquired $b=0$ s/mm² volume. Subsequently, the ROIs were propagated to the other DWI images in the scanned series. For analysis, the tonsils were regarded as a single organ. Voxel-wise fitting was performed with an in-house fitting script in MatLab 2017a (MathWorks, Natick, MA), which employs a variable starting point method before direct fitting of the bi-exponential with the most suitable starting point. The range in which starting points were chosen was the same as the ranges chosen for the optimization of f , D^* , and D . For S_o , the range was widened to 200 to 5000 and for K the upper bound was rounded to 3. Fitting constraints were set to prevent severe outliers, but aimed to avoid skewing the distribution at the edge of the physiologically plausible parameter values. The used constraints were S_o in [0, 10000] [arbitrary units], f in [-1, 1] [-], D^* in [0, 0.8] mm²/s, D in [0, 0.02] mm²/s, and K in [-5, 5] [-], as a compromise between unconstrained and more strictly constrained fitting.

Comparison of the optimized set of 20 b-values with the reference set

To assess the change in precision due to b-value optimization, the non-optimized reference set was compared with the optimized set of 20 b-values from the first optimized acquisition in terms of mean and SD of the four parameters of interest (f , D^* , D , K) within the ROI. The SD in the ROI consists of both underlying physiological differences and noise. Because the underlying physiological differences are constant, the SD within the ROI was used as a

measure of precision. Only the fully corrected data (with applied distortion correction, and intra- and inter-volume registration) were used in this comparison. A Wilcoxon signed-rank test was used to test significance for both metrics and all parameters, leading to a total of 8 tests.

Repeatability assessment for type of registration and set size

To assess the repeatability between the two identical optimized non-swallowing acquisitions for the different set sizes and registration scenarios, the coefficient of repeatability (CR) [79] was used. The third repetition contained swallowing artifacts and was therefore not used in this assessment.

$$CR = 2.77 \sqrt{\frac{1}{2n} \sum_j (\overline{p_{j,A}} - \overline{p_{j,B}})^2} \quad (2.6)$$

where $\overline{p_{j,k}}$ is the ROI mean of the parameter of interest for each subject j at time point k (either A or B) and n is the number of subjects. The workflow for obtaining the CR is schematically depicted in Figure 2.1. Then the relative CR (relCR) is defined as

$$relCR = \frac{CR}{\overline{P}} \cdot 100 \quad (2.7)$$

where \overline{P} is the overall ROI mean over the two acquisitions across all volunteers:

$$\overline{P} = \frac{1}{2n} \sum_{j,k} \overline{p_{j,k}} \quad (2.8)$$

Similarly, the relative difference in parameter mean is defined as:

$$relDifference = \frac{\overline{\overline{p_{j,A}} - \overline{p_{j,B}}}}{\overline{P}} \cdot 100 \quad (2.9)$$

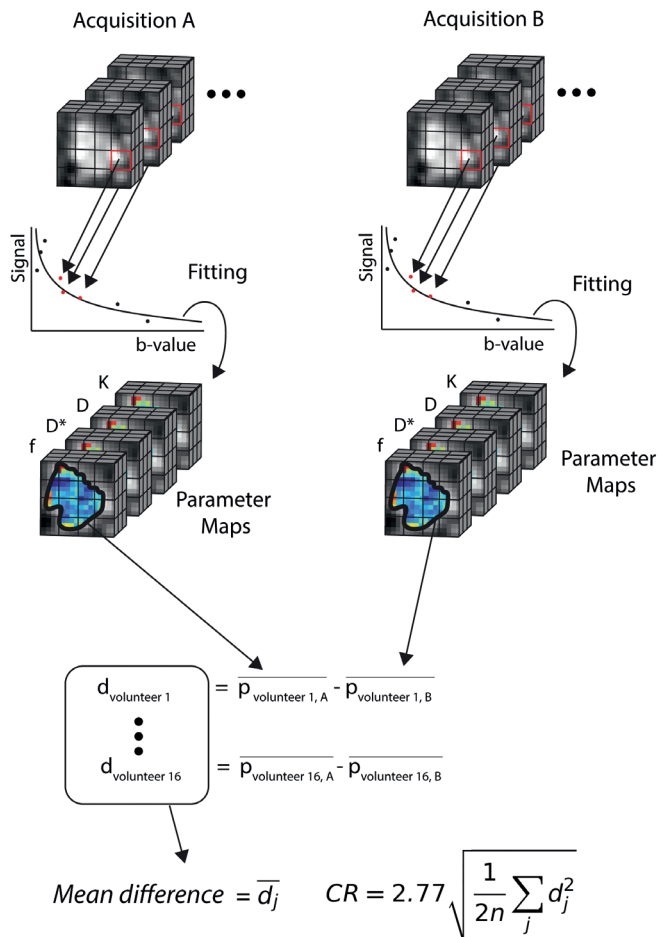


Figure 2.1. Overview of the workflow for obtaining the difference in mean and CR. Each acquisition (denoted A and B), contained multiple b-values. After distortion correction and registration, a fit can be made for each voxel, yielding estimates of the 4 IVIM-DKI parameters. Subsequently, the mean in the tonsils was calculated over both tonsils for each parameter for acquisitions A and B and the mean of acquisition A was subtracted from the mean of acquisition B to yield d_j . These steps were repeated for each volunteer. Finally, the average difference in mean and CR were calculated. CR, coefficient of repeatability; IVIM-DKI, intravoxel incoherent motion diffusion kurtosis imaging.

The optimized set of 30 b-values was compared to the sets of 20, 15, 10, 6 and 5 in terms of relCR and relDifference. For set size assessment, only the fully corrected data (applied distortion correction, intra- and inter-volume registration) were used. Wilcoxon signed-rank tests were applied to test if

relDifference was statistically significantly different from 0, for all set sizes and parameters, resulting in 24 tests.

For comparison of registration methods, only the optimized set size of 30 b-values was used. Comparisons in relDifference and relCR were carried out between the no registration and inter-volume only registration cases as well as between the no registration and intra- and inter-volume registration cases. Wilcoxon signed-rank tests were applied to test if the relDifference was different from 0 for all registration cases and all parameters, resulting in 12 tests.

Assessment of swallowing artifact rejection

In case of swallowing artifact image rejection, the ROI mean before rejection was compared to the ROI mean after rejection, and a Wilcoxon signed-rank test was applied. Comparisons were done for each set size applying intra- and inter-volume registration, resulting in 24 comparisons. Additionally, in case a significant difference ($p \leq 0.05$) was found before and after swallowing artifact rejection, the mean before and after rejection was compared to that of the second optimized acquisition without swallowing to assess improvement in accuracy, leading to another 18 comparisons.

2.3 Results

Figure 2.2 shows an example of the tonsil ROIs and corresponding parametric maps. One volunteer was excluded because the tonsils of this volunteer had been removed.

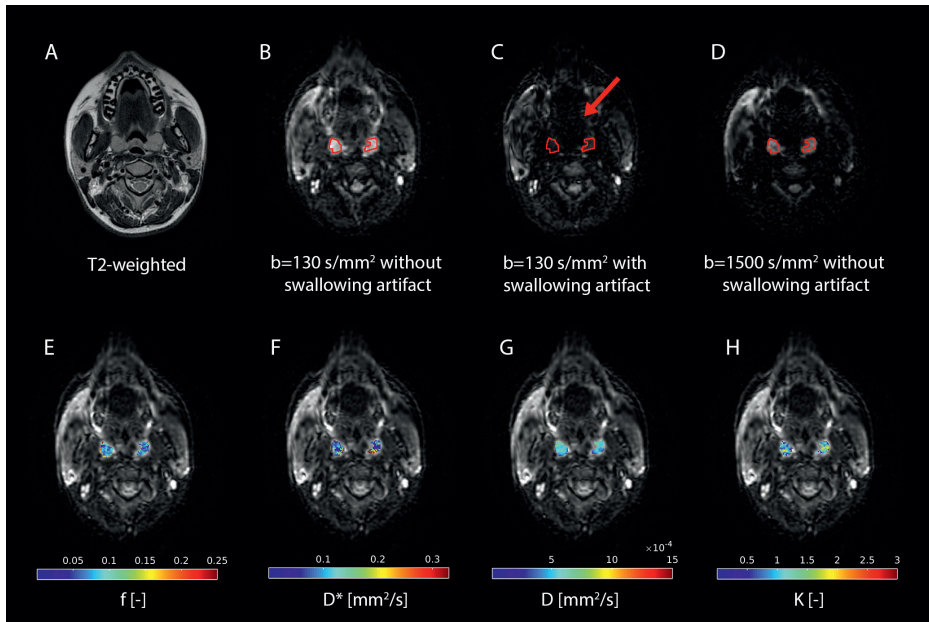


Figure 2.2. (A) Axial T2-weighted slice with (B) corresponding $b=130 \text{ s/mm}^2$ axial slice with the tonsils outlined in red and (C) identical slice to B but affected by a swallowing artifact. In (D) the same slice is shown but for $b=1500 \text{ s/mm}^2$. In the bottom row parametric maps for the tonsils are shown as color overlays on the corresponding $b=0 \text{ s/mm}^2$ slice with in (E) f , in (F) D^* in mm^2/s , in (G) D in mm^2/s , and in (H) K . D , diffusion coefficient, D^* , pseudo-diffusion coefficient; f , perfusion fraction; K , kurtosis.

2.3.1 Optimization of b-values

Table 2.3 shows the resulting optimized sets for 5, 6, 10, 20, and 30 b-values. Note that all b-values are rounded to multiples of ten to comply with the scanner's requirements. The increase in cost function value due to rounding was less than 1%, except for the set of 5 b-values where it was 1.8%. The mean and SD over all volunteers and set sizes for the two repeated fully registered non-swallowing acquisitions of the mean f , D^* , D , and K within the ROIs were 0.12 ± 0.06 [-] for f , $0.07 \pm 0.03 \text{ mm}^2/\text{s}$ for D^* , $0.8 \cdot 10^{-3} \pm 0.2 \cdot 10^{-3} \text{ mm}^2/\text{s}$ for D , and 0.73 ± 0.53 [-] for K . The obtained values of S_0 were 1021 ± 136 [a.u.].

Table 2.3. Rounded results of the optimization.

	b-values [s/mm²]
5 b-values	0, 20, 130, 780, 1500
6 b-values	0, 20, 80, 640, 790, 1500
10 b-values	0, 20, 2x80, 570, 2x770, 780, 2x1500
15 b-values	0, 10, 2x80, 130, 570, 2x770, 2x780, 790, 4x1500
20 b-values	0, 10, 3x80, 130, 570, 2x770, 2x780, 3x790, 6x1500
30 b-values	0, 2x10, 6x80, 280, 2x750, 760, 2x770, 2x780, 5x790, 8x1500

Comparing the relative CRLB from the reference set with the optimized set in the simulated ground-truth parameter sets, demonstrated that the reference set was expected to have a factor 2.2 higher variance than the optimized set for K . Slightly lower variance was expected in the reference set for f , D , and D^* with factors 0.5, 0.59, and 0.56, respectively, compared to the optimized set.

Comparison of the optimized set of 20 b-values with the reference set

The SD for D , K , and f was significantly lower in the optimized set of 20 b-values compared to the reference set that also had 20 b-values (D : $p=0.04$, K : $p=0.003$, f : $p=0.002$), as is shown in Figure 2.3. This corresponded to a 1.7, 1.5, and 2.3 times higher SD in the reference set in comparison to the optimized set in f , D , and K , respectively. The SD of D^* was significantly higher in the optimized set ($p=0.006$), corresponding with a 0.7 times lower SD in the reference set versus the optimized set. On average over all parameters, the improvement in SD was a factor 1.55. A statistically significant difference in mean was only seen for D^* ($p=0.01$). A non-physiological mean for the kurtosis ($K<0$) was found in 3 volunteers for the reference set, and in one volunteer for the optimized set of 20 b-values. Additionally, a non-physiological mean for f ($f<0$) was found in one volunteer in the reference set and in none of the volunteers in the optimized set of 20 b-values. Boxplots of the difference in mean of the optimized set versus the reference set can be found in Supporting Information Figure 2.S1.

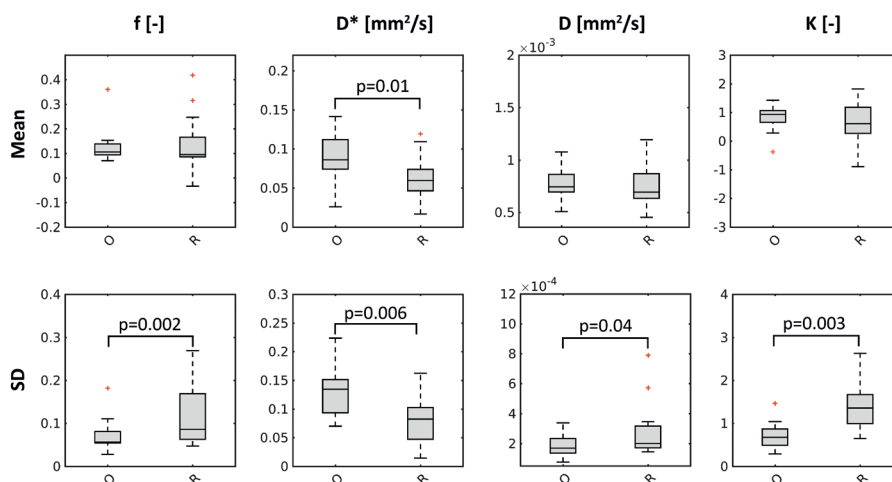


Figure 2.3. Boxplots showing the distribution of the mean and standard deviation (SD) in the tonsils over the volunteers for each of the four parameters of interest; f , D^* , D , and K . The optimized set of 20 b-values is marked 'O' and the reference set 'R'. p-values are indicated in case $p \leq 0.05$ according to the Wilcoxon signed-rank test.

The effect of the number of b-values on parameter precision

Figure 2.4 shows relDifference and relCR each for set size. The relCR of D was 37.7% for a set size of 30 and varied only slightly for set sizes of 20 and 15. However, for set sizes smaller than 15 b-values, a considerable increase in relCR was observed up to 73.7% for a set size of 6. The pattern in D^* was similar to D , with relCR ranging from 74.3% in the set size of 20 b-values up to 110.4% in the set size of 5 b-values. However, the relCR of D^* in set size 30 was 124.2%, and therefore notably higher than any of the other set sizes. For f , the relCR was lowest in set sizes 15 and 20 and slightly higher in the other set sizes. A slight decrease from 116.4% to 97.7% in relCR was seen in set size 20 compared to 30 in K . When removing more b-values from the set, relCR of K increased continuously up to 292.8% in the set of 5 b-values. Significant differences in parameter mean were found in the set size 10 for f (31.6%) and the set size 20 for D^* (23.0%). However, the relDifference of f was substantial with a range of 17.1% to 31.6% in all set sizes. Boxplots of the bias of set size 30 versus the other set sizes can be found in Supporting Information Figure 2.S2.

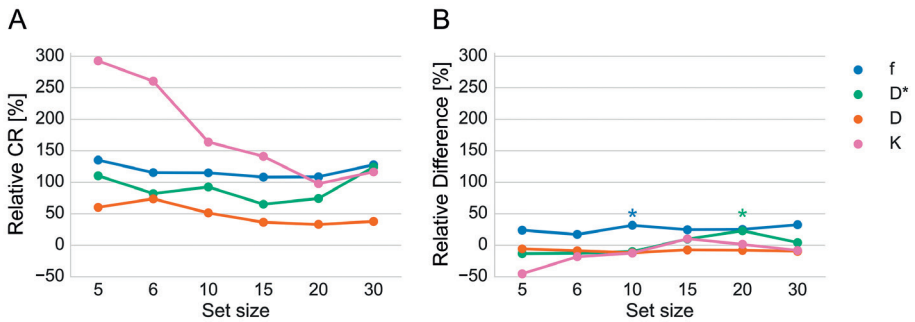


Figure 2.4. (A) Relative CR and (B) relative difference in mean between two identical acquisitions shown as a percentage of the overall mean of the two acquisitions for the four parameters of interest; f , D^* , D , and K . Significant differences in mean based on Wilcoxon signed-rank tests were found in f for set size 10 and D^* for set size 20. These points are denoted by an asterisk marker. For an explanation of the metrics, see Figure 2.1.

2.3.2 Registration

When only inter-volume registration was applied, correction of severe registration errors was necessary. No registration errors were found in case intra-volume registration was also applied. Figure 2.5 shows relDifference and relCR for each registration scenario. Inter-volume only registration showed a smaller relCR by 25 to 40% in f compared to the other scenarios. A similar effect was seen in D where the decrease was around 10%. However, the opposite effect was seen in D^* and K , where relCR increases of 1% to 23% and 30% to 42% were observed respectively. Significant differences in parameter mean were found for no registration (31.5%) and inter-volume only registration (22.2%) for f , leaving the full registration as the only scenario without a significant difference in parameter mean in any parameter. Nevertheless, the 32.5% relDifference for f of in the full registration scenario was still substantial. Boxplots of the difference in mean of the fully registered set versus the other two registration procedures can be found in Supporting Information Figure 2.S3. Parameter maps for each registration scenario and parameter are shown for one volunteer in Supporting Information Figure 2.S4.

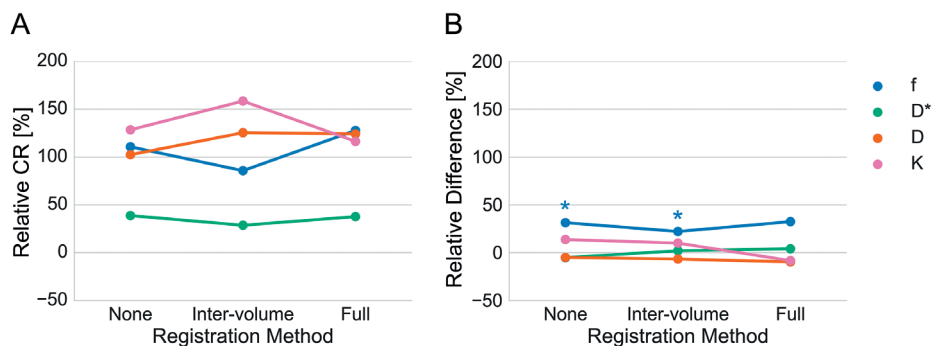


Figure 2.5. (A) Relative CR and (B) relative difference in mean between two identical acquisitions shown as a percentage of the overall mean of the two acquisitions for the four parameters of interest; f , D^* , D , and K . Full registration comprises intra- and inter-volume registration. Significant differences in mean based on Wilcoxon signed-rank tests were found in f for no registration and inter-volume only registration. These points are denoted by an asterisk marker. For an explanation of the metrics, see Figure 2.1.

2.3.3 Swallowing artifact rejection

Table 2.4 shows the average percentage of rejected b-value volumes per set size due to swallowing artifact presence. Significant differences in ROI mean between artifact-rejected and non-artifact-rejected data were found for a set size of 5 for all parameters (f : $p=0.001$, D^* : $p=0.007$, D : $p=0.0005$, K : $p=0.003$), as well as in set sizes 10 ($p=0.01$), 15 ($p=0.008$) and 20 ($p=0.008$) for f and set size 15 ($p=0.006$) and 20 ($p=0.03$) for D . In all mentioned cases, except D , a significant difference in ROI mean was observed between the non-rejected data and the second optimized acquisition. This difference was resolved after artifact image rejection, as is illustrated by Figure 2.6 for the mean perfusion fraction, indicating repeatability is improved after artifact image rejection. Boxplots for the other parameters can be found in Supporting Information Figure 2.S5. Supporting Information Figure 2.S6 contains scatter plots of the data before and after swallowing rejection plotted versus the baseline for all parameters and set sizes.

Table 2.4. Average percentage of rejected data per set size over 16 volunteers.

Set size	Average amount rejected b-values (%)
30	9.6
20	15.3
15	15.0
10	12.5
6	4.2
5	16.3

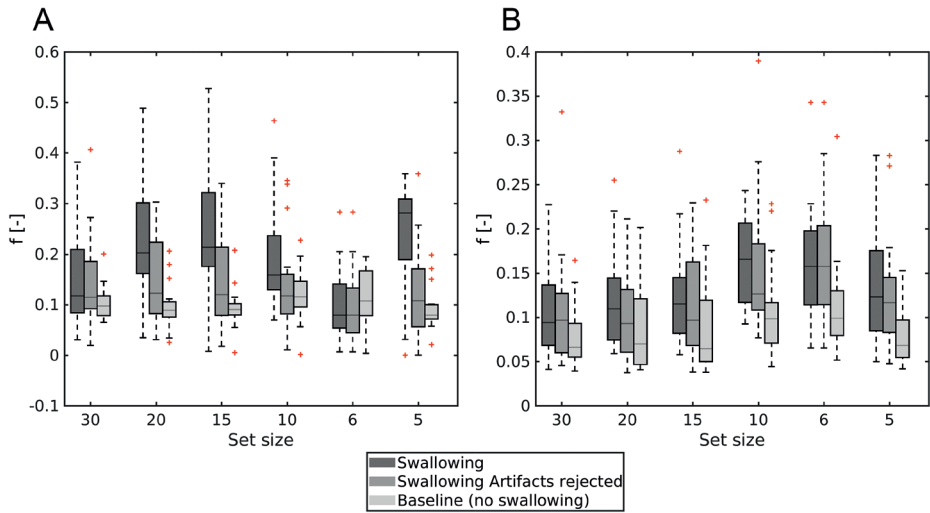


Figure 2.6. Boxplot of the mean (A) and standard deviation (B) of f over 16 volunteers for each set size of 5 to 30 b-values. The acquisition including swallowing artifacts (darkest gray) performs worse than the same acquisition after swallowing artifact correction (middle gray) when compared to the baseline acquisition (lightest gray) which did not contain any swallowing artifacts.

2.4 Discussion

Hybrid IVIM-DKI is promising for response assessment of head and neck cancer during (chemo)radiotherapy [42, 57]. However, inefficient sampling of b-values and motion corruption hamper precise parameter estimation in hybrid IVIM-DKI and thus diminish its potential for response assessment. In this paper, we showed that acquisition optimization and motion correction improves parameter estimation precision of the parameters of interest. Additionally, parameter repeatability metrics did not improve for sets larger than 15 b-values, suggesting that a set of 15 b-values is sufficient. Possibly because image acquisition noise is not the dominant error term for larger sets. No benefit of registration was found in our data, but swallowing artifact image rejection was beneficial, especially if a larger part (>10%) of the data is affected by such artifacts. To the best of our knowledge, this is the first study that addressed combined acquisition optimization and motion correction for hybrid IVIM-DKI in head and neck.

Most b-values in the optimal sets were near or above 800 s/mm^2 (restricted diffusion regime), followed by b-values below 200 s/mm^2 (perfusion regime), whereas relatively few b-values were chosen in between, in the free diffusion regime. Our findings suggest that the distribution of the b-values reflects the

uncertainty in the main parameters that rely on each regime. D was the most reliable parameter, followed by f and D^* . K relies mainly on high b -values and was the least reliable parameter.

Using a higher maximum b -value than 1500 s/mm^2 is expected to improve the precision of K , as long as the signal intensity at the maximum b -value is well above the noise floor to avoid parameter estimation bias. Since the model predicts an increase in signal at very high b -values, it is not expected that the optimization reaches a natural maximum b -value. Instead, it would venture into b -values where the kurtosis model is not valid. In this work, we chose a maximum b -value of $b=1500 \text{ s/mm}^2$ due to the limited SNR at this b -value in the head and neck region at 1.5 tesla to avoid the noise floor and the related parameter estimation bias.

Previous studies on b -value optimization were carried out for the mono-exponential DWI model [60–62] and the IVIM model [67, 68, 70]. These studies have largely reported an amount of unique b -values equal to the minimum of b -values needed to estimate all parameters. In this work, however, sets larger than 5 b -values always consisted of more than 5 unique b -values. Nonetheless, in the mono-exponential model, Brihuega-Moreno et al. [62] also reported more variation in b -values than the minimum needed for estimation; however, this was in the presence of large parameter ranges and without maximum b -value constraint. This could be in line with this study because the parameter ranges applied in this study were quite broad, and due to the inclusion of K , the b -value range was also larger. Additionally, a larger distribution was seen in more complex problems, for example, in optimizing gradient strengths and directions for diffusion kurtosis imaging [80]. Yet, even then discretization is seen.

The optimization technique discussed in this paper consists of a general framework and can be tailored to specific interests by weighting the parameters, or even choosing another model. We chose an equal weighting for all the parameters of interest (f , D^* , D , and K). However, if there is no interest in one specific parameter it can be left out of the optimization. Similarly, if one parameter is more important than the others, a higher weight can be assigned to this parameter. Therefore, specifying the objective function for b -value optimization should be adapted according to specific interests and is therefore subjective in nature [70].

Standardization of DWI across imaging systems is challenging [81]. The presented optimization framework depends on two parameters that are system dependent: SNR and TE. The b -value set is quite robust to SNR with only

about 1% variation in cost between the set optimized in this manuscript and the set optimized at an SNR of 15 to 25. Because the TE correction primarily affects simulated SNR, it is expected that the b-value set is still relatively optimal, even if the TE used and SNR achieved are slightly different across systems. Therefore, the optimal b-value sets that we propose could be of interest as a first step towards standardization.

Compared to the reference b-value set of 20 b-values, the optimal b-value set showed superior SD in f , D , and K , and slightly inferior SD in D^* . This indicates that precision is improved because the underlying physiological variation was assumed to remain constant within the scan session. Interestingly, based on the relative CRLB, a higher variation of a factor 2.2 for K was expected in the reference set compared to the optimized set, and a slightly decreased variation in f , D , and D^* (factor 0.5, 0.59, and 0.56, respectively) in the reference set compared to the optimized set. Thus, the actual optimized set performed better than expected for D and f , showing there are more factors at play than the image noise described by the CRLB. Repeatability could not be assessed because a repeat of the reference scan was not acquired due to scan time limitations.

The repeatability of all parameters seems worse for the set of 30 b-values than for the set of 20 b-values. This seems counterintuitive but is likely due to the longer scan time, resulting in an opportunity for motion and other effects corrupting the data that could eventually lead to worsening the parameter estimation. The registration method we applied was not sufficient to account for all the time-related effects. Also, no clear impairment in repeatability metrics was observed for any of the parameters when moving from a set of 20 to a set of 15 b-values. A trade-off between scan time and parameter repeatability was only observed when imaging with less than 15 b-values.

Repeatability is expected to improve in patients, due to higher SNR often encountered in tumors than in healthy tissue for diffusion imaging. This is because healthy tissue has a lower intravoxel water content than non-necrotic tumor tissue. Therefore, even if the measurement variation in f , D^* , and K is high, it might become acceptable in patients. Especially K is expected to improve because it mainly relies on high b-values where the increase in SNR has the highest impact.

No clear advantage is shown for applying inter- and intra-volume registration in this work. This might be because volunteers generally lie very still, which would not necessarily be the case in patients. It might also be that once the estimation is based on a multi b-value fitting, a sufficient number of b-values

makes the estimation robust enough regarding motion. However, a side effect of intra-volume registration is smoothing of the b-value images. This could lead to smoothing of underlying physiological differences and might be problematic when assessing intra-tumor heterogeneity.

In this study, we have shown that swallowing artifact image rejection improves the accuracy of parameter estimation. Previously, Chevallier et al. [82] already showed that rejection of motion-corrupted volumes is beneficial for the estimation of IVIM parameters. In this paper, the clearest effect is seen in the perfusion fraction, especially in set sizes 20, 15, 10, and 5. This could be because the relative signal drop is larger in low b-values ($b < 200 \text{ s/mm}^2$) than in higher b-values. Furthermore, our results suggest that there is a threshold for the amount of artifacts of around 10% that needs to occur before swallowing artifact image rejection is worthwhile. In set sizes 6 and 30 no strong swallowing artifact image rejection effect was observed, whereas in a set size of 5, 10, 15, and 20 more than 10% of data was rejected and a clear effect is seen in the perfusion fraction.

Because swallowing artifacts can severely hamper parameter estimation, swallowing artifact rejection is beneficial. An automated strategy for swallowing artifact rejection, such as proposed by Gurney-Champion et al. [83], could be a relatively simple way to implement this procedure in clinical practice. As a topic for further research, the effect of swallowing artifact correction in the head and neck region could be investigated, for example, using a combined principal component analysis and temporal maximum intensity projection approach PCATMIP [84].

The main limitation of this work is that it was performed on healthy volunteers, for whom the SNR at high b-values was relatively low compared to tumor tissue. Another limitation is that a correction for multiple comparisons was not applied. For these two reasons, the results should be independently validated in a larger cohort of head and neck cancer patients.

2.5 Conclusion

The effect of b-value optimization, protocol efficiency, registration, and swallowing artifact image rejection on parameter precision of hybrid IVIM-DKI was assessed. Optimization of b-values is recommended because it improved the precision of 3 (D , f , K) out of 4 parameters of interest compared to the reference set. The b-value set of 15 images ($b=1x0$, $1x10$, $2x80$, $1x130$, $1x570$, $2x770$, $2x780$, $1x790$ and $4x1500 \text{ s/mm}^2$) yielded the optimal tradeoff between scan time and parameter precision, with a repeatability comparable to the

set of 30 in half the scan time. No clear advantage of image registration was demonstrated. However, swallowing artifact image rejection was beneficial when more than roughly 10% of the images contained artifacts.

2.6 Supporting information

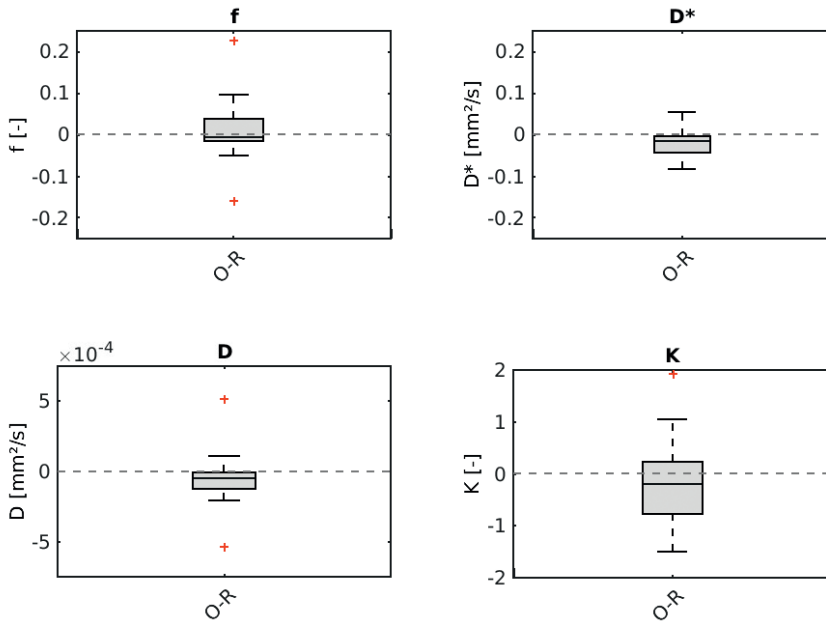


Figure 2.S1. Difference in mean between the optimized set of 20 b-values "O" and the reference set "R" for the four parameters of interest; the perfusion fraction f , the pseudo-diffusion coefficient D^* , the diffusion coefficient D , and the kurtosis K .

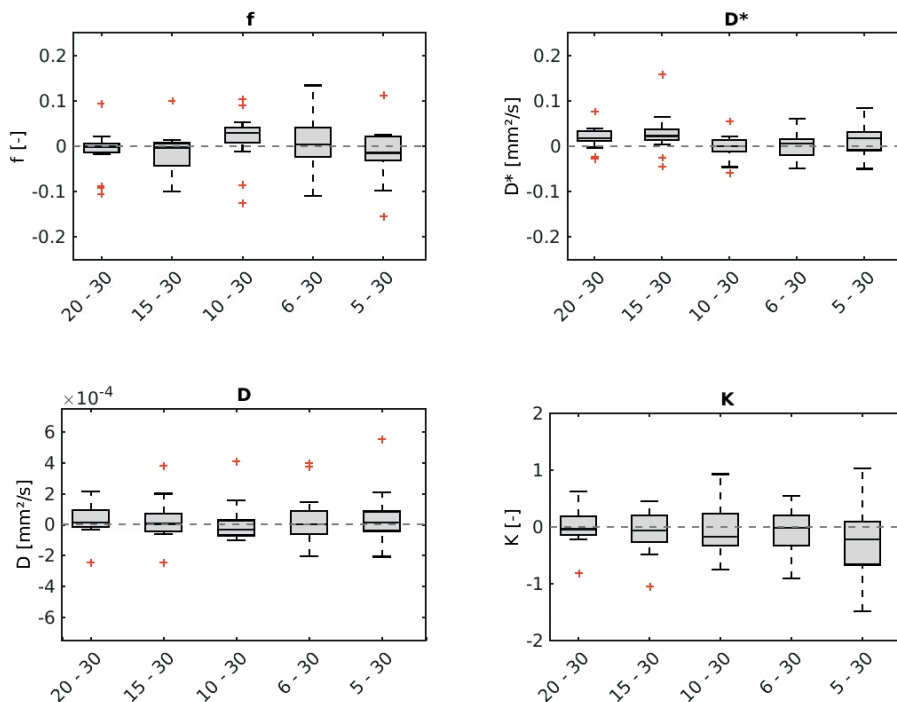


Figure 2.S2. Difference in mean between the indicated set sizes for each parameter and the set of 30 b-values for the four parameters of interest; the perfusion fraction f , the pseudo-diffusion coefficient D^* , the diffusion coefficient D , and the kurtosis K .

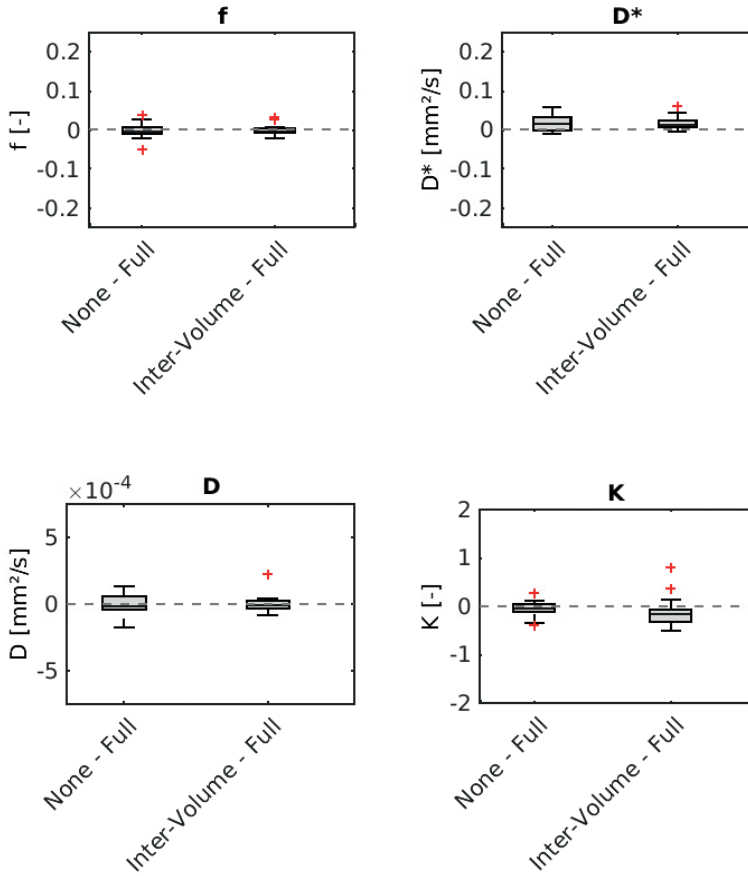


Figure 2.S3. Difference in mean between full registration (intra- and inter-volume registration) and indicated registration scenarios for the four parameters of interest; the perfusion fraction f , the pseudo-diffusion coefficient D^* , the diffusion coefficient D , and the kurtosis K .

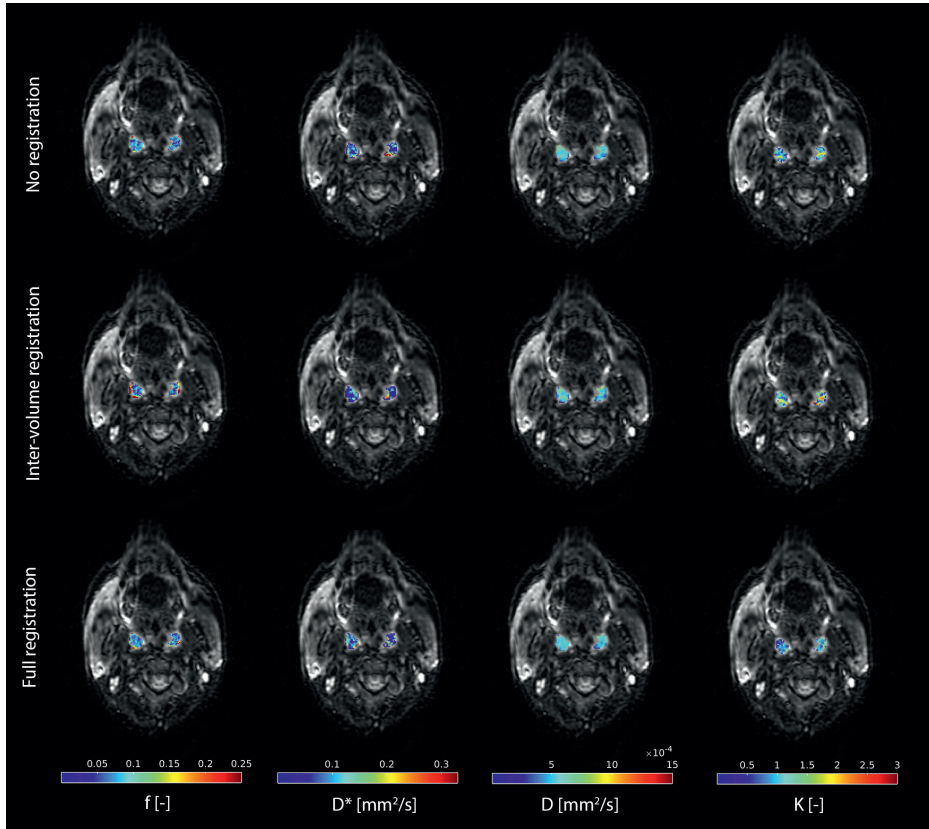


Figure 2.S4. Example parameter maps for each registration scenario (top to bottom). Full registration includes intra- and inter-volume registration. Distortion correction was applied to all scans before registration. From left to right the four parameters of interest are depicted; the perfusion fraction f , the pseudo-diffusion coefficient D^* , the diffusion coefficient D , and the kurtosis K .

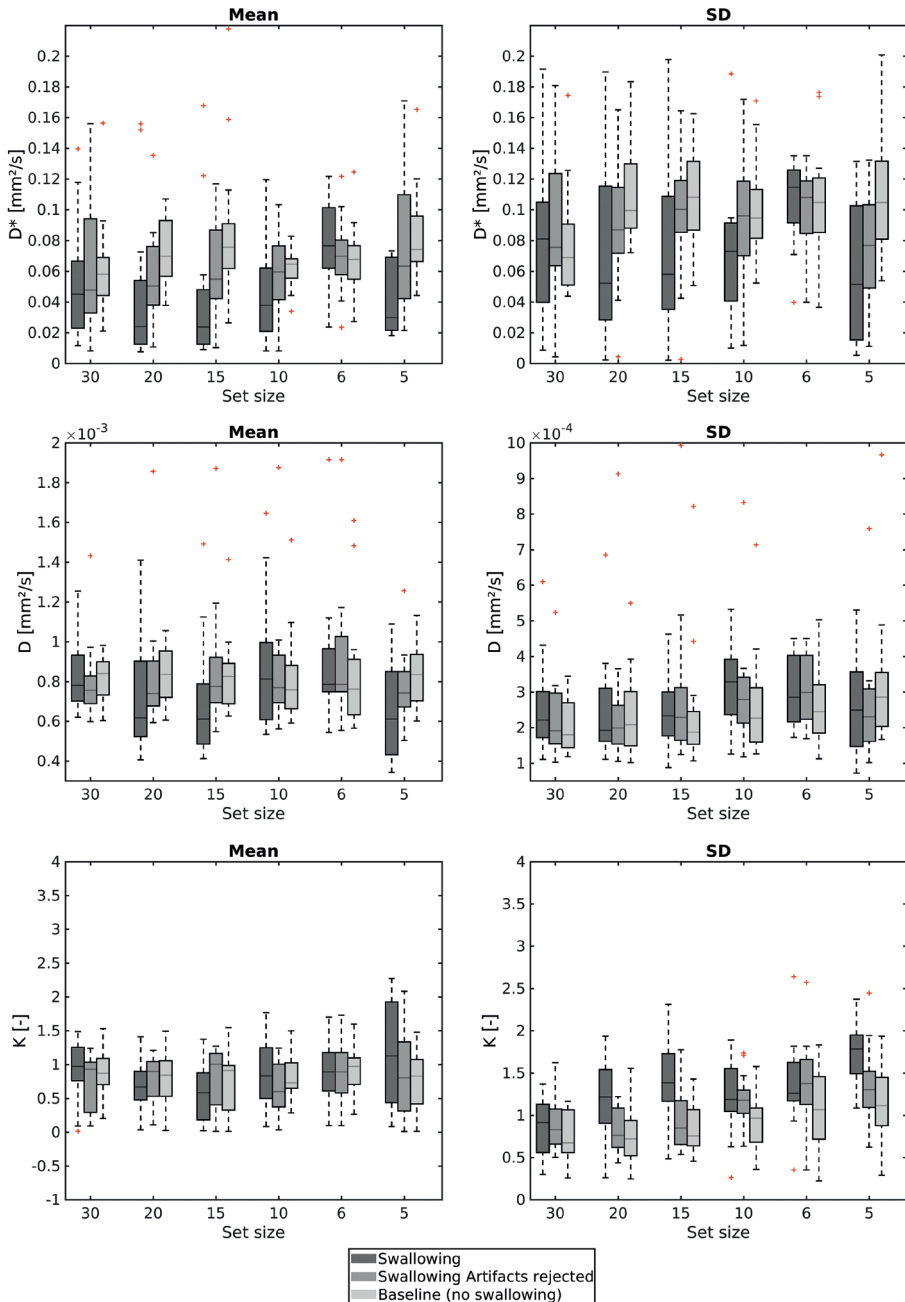


Figure 2.S5. Distributions of mean and the standard deviation (SD) for acquisitions with swallowing, with swallowing artifact images rejected, and without swallowing for the pseudo-diffusion coefficient D^* , the diffusion coefficient D , and the kurtosis K (top to bottom).

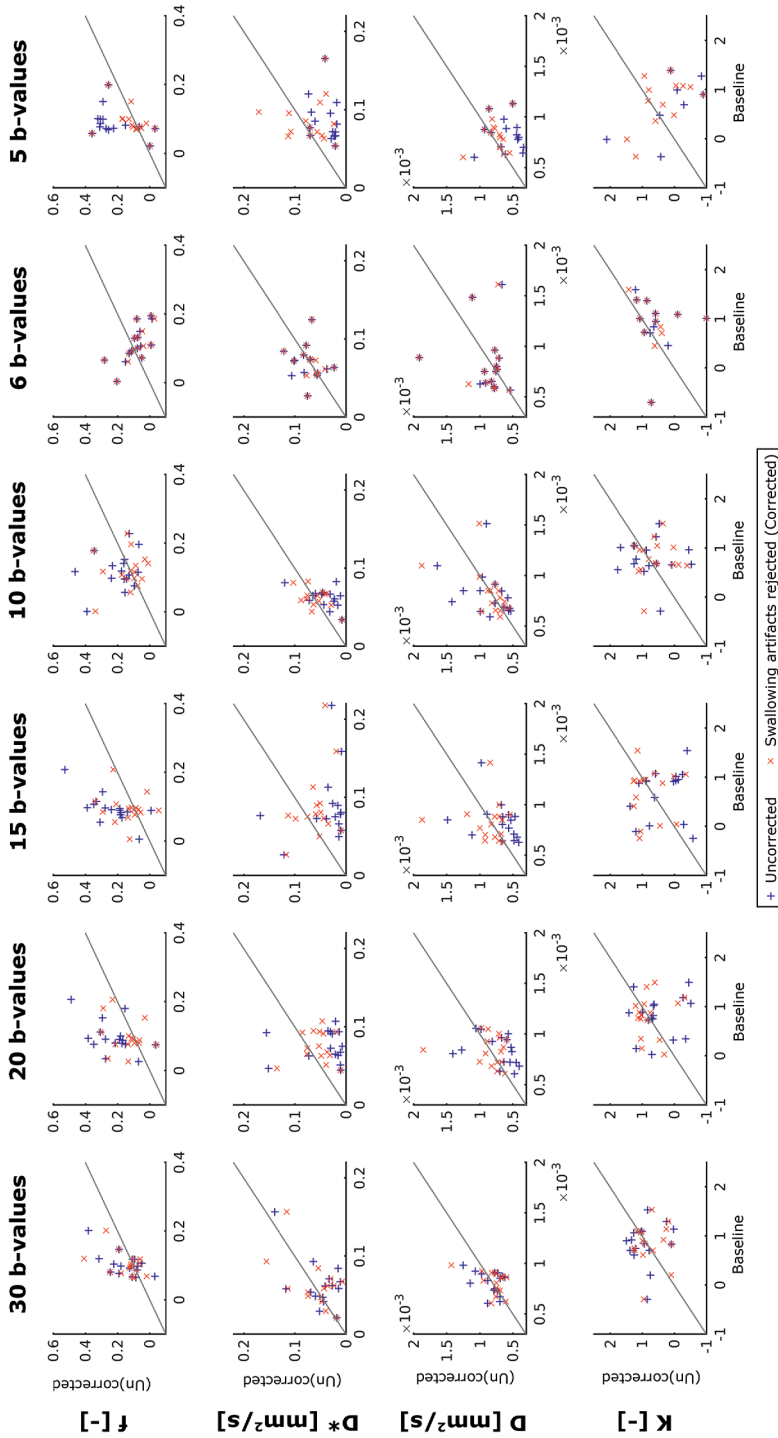


Figure 2.S6. Comparison of mean values before (blue) and after (red) swallowing artifact rejection versus the baseline mean without swallowing for each of the volunteers. From top to bottom the four parameters of interest; the perfusion fraction f , the pseudo-diffusion coefficient D^* , the diffusion coefficient D , and the kurtosis K . From left to right the size of the b-value set ranging from 30 to 5 b-values. The gray line indicates the unity line.

3

Multi-delay pseudo-continuous arterial spin labeling to measure blood flow in the head and neck

Nienke D. Sijtsma, Steven F. Petit,
Gerda M. Verduijn, Dirk H.J. Poot,
Esther A.H. Warnert, Mischa S. Hoogeman,
Juan A. Hernandez-Tamames

This chapter is based on Sijtsma ND, Petit SF, Verduijn GM, Poot DHJ, Warnert EAH, Hoogeman MS, et al. Multidelay pseudocontinuous arterial spin labeling to measure blood flow in the head and neck. NMR Biomed. 2023;36:e4898.

Abstract

Perfusion MRI is promising for the assessment, prediction, and monitoring of tumor response and radiation toxicity in organs at risk in head and neck cancer. Arterial spin labeling (ASL) may be an attractive alternative to conventional perfusion MRI, that does not require the administration of contrast agents. However, currently, little is known about the characteristics and performance of ASL in healthy tissues in the head and neck region. Therefore, the purpose of this study was to optimize and evaluate multi-delay pseudo-continuous ASL (pCASL) for the head and neck region and to explore nominal values and measurement repeatability for the blood flow (BF), and the transit time and T1 values needed for BF quantification in healthy tissues. Twenty healthy volunteers underwent a scan session consisting of four repeats of multi-delay pCASL (post-label delays: 1000, 1632, 2479 ms). Regions of interest were defined in the parotid glands, submandibular glands, tonsils, and the cerebellum (as a reference). Nominal values of BF were calculated as the average over four repeats per volunteer. The repeatability coefficient and within-subject coefficient of repeatability (wCV) of BF were calculated. The effect of T1 (map vs. cohort average) and transit time correction on BF was investigated. The mean BF (\pm SE) was 55.7 ± 3.1 ml/100 g /min for the parotid glands, 41.2 ± 2.8 ml/100 g/min for the submandibular glands, and 32.3 ± 2.2 ml/100 g/min for the tonsils. The best repeatability was found in the parotid glands (wCV=13.3%-16.1%), followed by the submandibular glands and tonsils (wCV=20.0%-24.6%). On average, the effect of T1 and transit time correction on BF was limited, although substantial bias occurred in individual acquisitions. In conclusion, we demonstrated the feasibility of BF measurements in the head and neck region using multi-delay pCASL and reported on nominal BF values, BF repeatability, the effect of T1, and transit time in various tissues in the head and neck region.

3.1 Introduction

Perfusion phenomena in tissues can be assessed by a number of techniques, including dynamic contrast enhanced (DCE) MRI, dynamic susceptibility contrast (DSC) MRI, and arterial spin labeling (ASL). Yet, while DCE and DSC have found their way into state-of-the-art imaging protocols for the head and neck region [85], ASL has only recently gained more attention for the head and neck region [86]. The advantage of ASL is that, in contrast to DCE and DSC, it does not require the injection of an external contrast agent. Recently, the potential value of ASL in head and neck cancer has been demonstrated for the detection of residual tumor after treatment [51], differentiating between tumor types [87, 88], determining tumor stage [89], differentiation between benign and malignant tissue [90], differentiation between metastatic and reactive lymph nodes [91], and even early treatment response [92]. In general, a good correlation between blood flow (BF) values measured by DCE and ASL [93, 94] is reported.

However, to date, the overwhelming majority of research on ASL in the head and neck region has focused on the tumor. Besides the tumor, perfusion MRI may have added value in the assessment, prediction, and monitoring of toxicity of the (chemo)radiotherapy treatment in important organs at risk in the head and neck. For example, changes in perfusion during treatment have been observed with DCE in submandibular and parotid glands [95, 96]. Avoiding the administration of a contrast agent, especially when sessions are repeated during treatment, would be preferable, hence the interest in ASL for normal tissues in the head and neck region.

Currently, ASL for the head and neck is under development, and very little is known about BF values and T1 in healthy tissues in the head and neck region, transit times of the blood to these tissues, and the repeatability of the BF values. Moreover, normative values are important to help put BF values and BF changes in patients into perspective; they also provide essential information for the design of (clinical) studies to analyze the normal tissue response to treatment and the possible incorporation of ASL into clinical protocols in the future.

Therefore, the aim of this study was to optimize and evaluate a multi-delay pseudo-continuous ASL (pCASL) sequence and explore nominal BF values for several tissues in the head and neck region. To this end, we investigated the repeatability of BF, obtained T1 values of the tissues, and investigated the influence of transit time correction and T1 value on BF measurements.

3.2 Materials and methods

3.2.1 Subjects

In total, 20 healthy subjects were included (mean age 25 years, age range 20–31 years, nine males, 11 females). This study was approved by our institutional medical ethical board (protocol 2014–096) and written informed consent was obtained from all subjects. Exclusion criteria were any metal-containing implants, fillings, or braces, in or around the oral cavity.

3.2.2 Imaging protocol

All subjects were scanned on a 3T MR750 Discovery system (GE Healthcare, Waukesha, WI, USA) with a 12-channel head and neck coil (GE Healthcare, Waukesha, WI, USA). The scan protocol consisted of a 3D multi-parametric quantification using an interleaved Look-Locker acquisition sequence with a T2 preparation pulse (3D-QALAS), three-delay encoded ASL, and multi b-value diffusion-weighted imaging. The 3D-QALAS [97] (0.5x0.5x2 mm³ voxels, scan time 4 min 4 s) scan was acquired in the axial plane. The SyMRI version 0.45.2721Q1 software package (SyntheticMR AB, Linköping, Sweden) was used to generate synthetic T1-weighted, T2-weighted and proton density weighted images, and a T1 map that was used for quantification of the ASL BF. Next, a 3D encoded pCASL scan was acquired that incorporated three post-label delays (PLDs) in one acquisition [98, 99] (3D stack of spirals fast spin echo with six arms, PLDs: 1000, 1632, 2479 ms, total Label Duration (LD): 3000 ms, Effective LD per PLD: 632 ms (PLD=1000 ms), 847 ms (PLD=1632 ms), 1521 ms (PLD=2479 ms), 2.2x2.2x4 mm³ voxels, 128x128x26 matrix (acquired resolution), TE=10.5 ms, TR=6012 ms, number of averages=2, scan time 5–7 min). Background suppression was implemented according to Maleki et al. [100]. The labeling plane was automatically placed 2 cm below the most caudal slice, which was positioned on the chin, to ensure the labeling plane was situated in the neck area (Figure 3.1). All ASL scans were acquired in the axial plane, and the subjects were positioned in the scanner as straight as possible (i.e., left and right symmetry in the axial plane), such that the labeling plane was, by approximation, perpendicular to the BF direction for all volunteers. Padding was used to minimize involuntary motion during the scan session. A three-delay pCASL was chosen as this is a robust protocol that would be suitable for a variety of tissues in the head and neck region. The three-delay pCASL scan was repeated 4 times while the subject remained in the scanner, with intervals of 5 minutes. The total scan time for each subject was 1 hour.

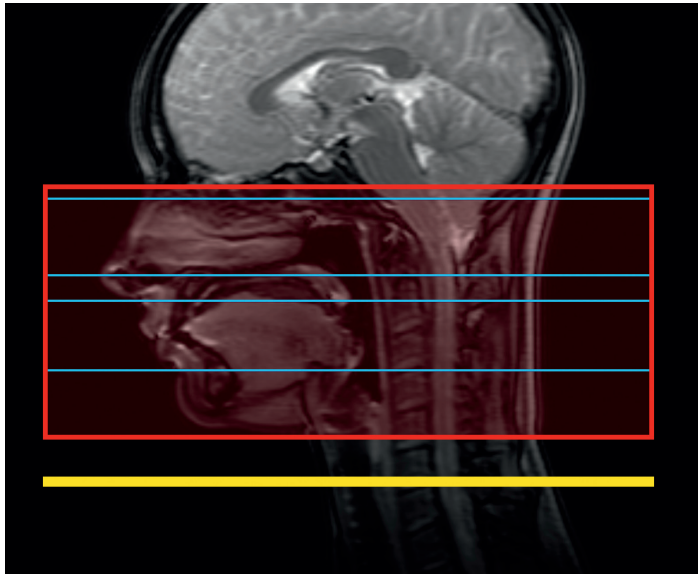


Figure 3.1. Example of the placement of the field of view (red) and labeling plane (yellow) for one volunteer, the blue lines indicate the slices as displayed in Figure 3.2. The labeling plane was automatically placed 2 cm below the most caudal slice of the field of view.

3.2.3 Post-processing

Organs of interest

The study focused on the parotid glands and submandibular glands because these are important organs at risk in head and neck radiotherapy treatments. In addition, the tonsils were added to represent a surrogate tumor location, to assess the feasibility of measuring BF in healthy tissue and tumor simultaneously in one comprehensive ASL protocol. All tissues of interest were delineated by an experienced radiation oncologist on the synthetic T2-weighted images obtained from 3D-QALAS. Because ASL in the brain has been described extensively, a region of interest (ROI) was also drawn in the cerebellum as a reference measurement to compare with previous literature. Each ASL scan was rigidly registered to the synthetic proton density weighted image from 3D-QALAS, because the proton density weighted contrast is most similar to the 2-second saturation recovery image from the ASL sequence. Subsequently, delineations were transferred to each ASL acquisition. To diminish the influence of delineation errors, registration errors, and the partial volume effect, so that only the tissue of interest is contained within the contours, the 3D delineated volumes of the parotid glands, submandibular glands, and tonsils were eroded by 2 mm in the axial plane and by 4 mm in slice direction. The eroded contours

are referred to as the ROIs. The cerebellum ROI was not eroded, because the cerebellum was not delineated, but an ROI was drawn inside the cerebellum instead.

ASL BF quantification

To avoid bias because of an imperfectly chosen PLD, the BF was transit time corrected through voxel-wise fitting of the ASL signal to Equation 3.1 using the three PLDs [98]. Fitting was performed for the ROI of each tissue of interest ROI separately.

$$BF = \frac{6000\lambda e^{\delta/T1_a}}{2\varepsilon T1_t \left(e^{-\frac{\max(PLD-\delta,0)}{T1_t}} - e^{-\frac{\max(LD+PLD-\delta,0)}{T1_t}} \right)} \frac{SI_{PW}}{\frac{SI_{SR}}{S_c}} \quad (3.1)$$

where SI_{PW} is the signal intensity of the perfusion weighted image and SI_{SR} is the signal intensity in the 2-second saturation recovery image of the ASL acquisition, which is divided by a factor $S_c = 1 - e^{-\frac{2}{T1_t}}$ to correct for partial signal recovery. δ , the arterial transit time, is estimated by the signal-weighted delay method as described by Dai et al. [98]. λ is the blood-tissue partition coefficient which was set to 0.9 g/ml. ε is a correction factor to account for the combined effect of labeling efficiency and background suppression and was set to 0.6 to conform manufacturer's guidelines. $T1_a$ is the T1 of arterial blood and was set to 1650 ms [50, 101]. $T1_t$ is the T1 of the tissue of interest and, unless specifically stated otherwise, was incorporated voxel-wise based on the T1 map obtained from QALAS.

3.2.4 Assessment

Nominal values of BF

Although BF values are by definition positive, the BF estimated by Equation 3.1 can be negative due to noise or motion causing misalignment between the control and labeled images. In the ROIs, the voxels containing a negative BF were excluded. Therefore, the mean BF was determined in each repeated acquisition only over those voxels containing a positive BF. Next, the mean BF per subject ($\overline{BF_S}$) was calculated per tissue by averaging over the four repeats. The nominal values of the BF in the tissues were defined by $\overline{BF_n} \pm SD$, where $\overline{BF_n}$ is the grand mean over the $\overline{BF_S}$ for the left and right side combined of each tissue for all subjects (i.e., 40 measurements in total per tissue) and its corresponding standard deviation (SD) and standard error (SE). The range of ROI means of all subjects and repeats is reported for each tissue of interest, as well as plotted per subject. Because T1 values of the tissues were necessary for the quantification of BF, the range of T1 values of the tissues is reported.

The \overline{BF}_S and T1 were compared between the parotid glands, submandibular glands, and the tonsils (left and right averaged) using a paired Student's t-test in R statistical software (version 4.1.2, R Foundation for Statistical Computing, Vienna, Austria). Because of the limited number of comparisons (six in total), p-values were not corrected for multiple comparisons and a p-value less than 0.05 was considered statistically significant.

Repeatability

To calculate the repeatability measures, the within-subject SD (wSD) of the ROI mean was estimated for each tissue of interest by fitting a linear mixed effect model where the ROI means (one per repeat per ROI) were grouped by subject. For this purpose, the lme4 package (version 1.1-27.1) [102] in R statistical software (version 4.1.2, R Foundation for Statistical Computing, Vienna, Austria) was used. Next, the repeatability coefficient, which indicates the minimum true difference, was estimated (\widehat{RC}) [103] by

$$\widehat{RC} = 1.96 \cdot \sqrt{2} \cdot wSD \approx 2.77wSD \quad (3.2)$$

The within-subject coefficient of variation was calculated as well, as this relative measure eases the comparison of repeatability between tissues:

$$wCV = \frac{wSD}{\overline{BF}_T} \quad (3.3)$$

where \overline{BF}_T is the nominal BF per tissue, left and right separated.

Influence of T1 and PLD

The average T1 per tissue over all volunteers was calculated to investigate the influence of T1 on the BF measurement, and to assess whether the acquisition of a T1 map could be omitted by using an average T1 in the BF estimation. For the parotid glands, submandibular glands, and tonsils, the left and right measurements were grouped before calculation of the average, so the average T1 was based on 40 ROI means for these tissues. A new BF map was constructed for each tissue according to Equation 3.1, with the exception that the $T1_T$ was set to the group average per tissue before voxel-wise fitting of the ASL signal, instead of using the T1 map. The difference between the BF based on the fixed T1 and based on the T1 map was calculated for each acquisition, plotted, and reported, to assess the possible bias in BF estimation if a T1 map is not available. Additionally, the wCV and RC were calculated for the BF based on a fixed T1 and compared with the wCV and RC calculated for the T1 map-based BF.

To investigate the possible underestimation of BF when using only one PLD, the BF was calculated voxel-wise according to the following equation for each PLD separately without transit time correction:

$$BF = \frac{6000\lambda e^{PLD/T1_a} S_{I_{PW}}}{2\varepsilon T1_t \left(1 - e^{-\frac{LD}{Tt}}\right) \frac{S_{I_{SR}}}{S_c}} \quad (3.4)$$

The difference between the three non-transit time corrected BF based on each of the three single PLDs and the transit time corrected BF values was calculated for each pCASL acquisition, then plotted and reported.

3.3 Results

Out of 20 volunteers, five volunteers had undergone resection of their tonsils and one presented only the right tonsil on the synthetic T1w and T2w scans. Therefore, the right tonsil was assessed in 15 volunteers and the left tonsil was assessed in 14 volunteers in total. Additionally, one repeat scan from one volunteer was fully excluded because of specific absorption rate (SAR) limitations. The size of the ROIs (i.e., after erosion in the case of the parotid glands, submandibular glands and tonsils) (\pm SD) was 6.8 ± 3.1 cc (range 2.0–15.9 cc) for the parotid glands, 1.9 ± 0.7 cc (range 0.5–3.3 cc) for the submandibular glands, 0.5 ± 0.5 cc (range 0.1–1.7 cc) for the tonsils, and 11.6 ± 4.2 (range 5.7–21.9 cc) for the cerebellum. The percentage of negative voxels that were excluded was 0.4% for the left parotid gland, 2.1% for the right parotid gland, 2.1% for the left submandibular gland, 3.8% for the right submandibular gland, 0.1% for both tonsils and 0.7% for the cerebellum.

Figure 3.2 shows an example of the delineations (before erosion) of the tissues in the head and neck region in several slices on the synthetic T2w image, T1 map, and BF map. The average T1 in the ROI ranged from 833 to 1645 ms for the parotid glands, 1079 to 1876 for the submandibular glands, and 1955 to 2892 ms for the tonsils, and was significantly different for these three tissues (Figure 3.3A). Table 3.1 shows the nominal values for the BF in the tissues in the head and neck, with their SD (which reflects the between-subject variability) and SE. BF was significantly different between the parotid glands, submandibular glands, and the tonsils (Figure 3.3B). As can be seen in Figure 3.3B, the ranges of BF in the tissues on the left and on the right generally coincided well. The mean BF for the right parotid gland ranged from 33.9 to 87.2 ml/100 g/min, while for the left parotid gland it ranged from 36.1 to 90.0 ml/100 g/min. For the right submandibular gland, the range was 20.6 to 61.6 ml/100 g/min, while for the left it was 21.0 to 80.6 ml/100 g/min. For the right tonsil, the range was

15.8 to 44.9 ml/100 g/min, while for the left it was 14.8 to 45.8 ml/100 g/min. However, the median BF for the left tonsil was higher than that for the right tonsil.

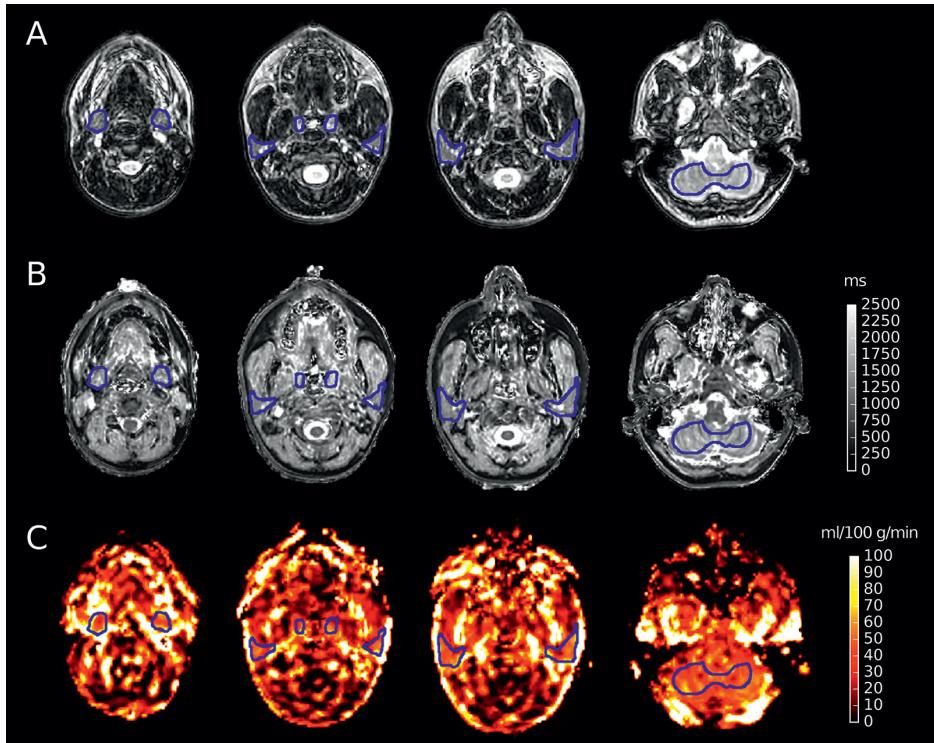


Figure 3.2. Example of the (A) T2w image, (B) T1 map, and (C) BF map for four different slices in one volunteer. Delineations (i.e., before erosion) are shown for the submandibular glands (first column), the tonsils (the two medial ROIs in the second column), and the parotid glands (third column). The cerebellum ROI is shown in the last column. The location of each of the slices is also indicated by blue lines in Figure 3.1. Note that delineations of the parotid glands are also visible in the second column. BF, blood flow; ROI, region of interest.

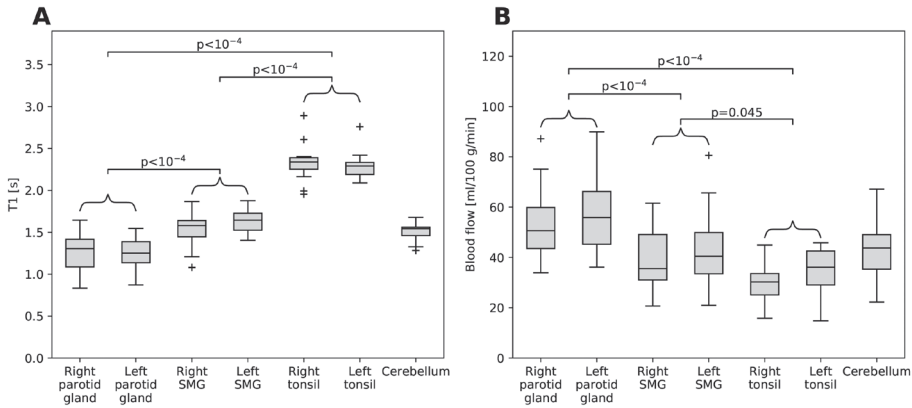


Figure 3.3. (A) Boxplot of the mean T1 for each tissue of interest per volunteer. Tonsils have a notably higher T1 than the other tissues, probably caused by high water content. Additionally, the range of T1 in the parotid glands is notably larger than in the other tissues. (B) Boxplot of the mean BF per volunteer (mean over the repeats) for each of the tissues of interest. The parotid glands showed the highest perfusion, while the tonsils displayed the lowest perfusion. For both (A) and (B), a paired t-test with the tonsils was performed with $n=14$ because some volunteers did not present with both tonsils, while comparisons between the parotid and submandibular glands were performed with $n=20$. BF, blood flow; SMG, submandibular gland.

Table 3.1. Nominal values of the BF with the corresponding SD and SE. BF, blood flow; SD, standard deviation; SE, standard error.

Tissue	BF [ml/100 g/min]	SD	SE
Parotid glands (n=20)	55.7	14.0	3.1
Submandibular glands (n=20)	41.2	12.4	2.8
Tonsils (n=14)	32.3	8.1	2.2
Cerebellum (n=20)	43.0	10.7	2.4

In general, all the tissues in the head and neck displayed similar transit times, on average (\pm SD), 1177 ± 76 ms for the right parotid gland, 1179 ± 66 ms for the left parotid gland, 1287 ± 165 for the right submandibular gland, 1204 ± 77 for the left submandibular gland, 1204 ± 128 ms for the right tonsil, and 1205 ± 119 ms for the left tonsil. However, the transit time for the cerebellum was notably longer, on average (\pm SD) 1715 ± 170 ms. This is graphically depicted in Figure 3.S1.

Figure 3.4 shows plots of the mean BF per repeat per volunteer in the parotid glands, submandibular glands, tonsils, and cerebellum. The majority of the ranges overlapped for the left and right sides for all tissues. However,

no overlap of the range of measurements between the left and right sides was present in two volunteers in the parotid glands, in two volunteers in the submandibular glands, and in three volunteers in the tonsils.

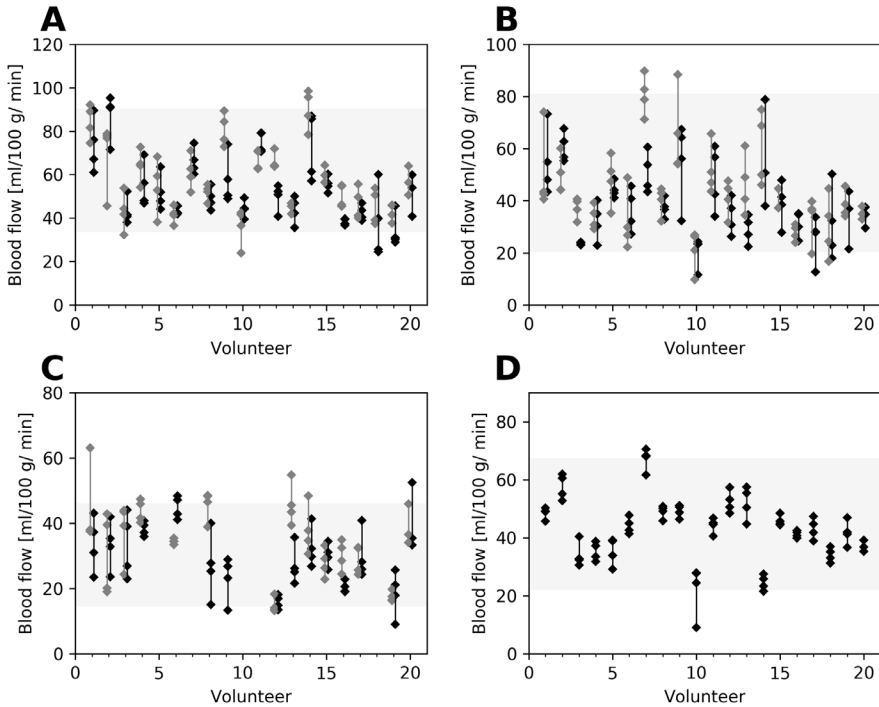


Figure 3.4. Plot showing the BF measurement for each acquisition for (A) the parotid glands, (B) the submandibular glands, (C) the tonsils, and (D) the cerebellum. The left side is denoted in gray and the right side in black (if applicable). The gray box indicates the minimum and maximum values of the mean BF per volunteer (shown as the extent of the boxplot in Figure 3.3B), with the left and right combined (if applicable). BF, blood flow.

Table 3.2 shows the repeatability of measurements. The best relative repeatability in terms of wCV was obtained in the cerebellum (9.0%), followed by the parotid glands (13.3%-16.1%). The tonsils and submandibular glands showed similar repeatability (20.0%-24.6%).

Table 3.2. Repeatability measures of BF for the assessed tissues. BF, blood flow; CI, confidence interval; RC, Repeatability coefficient; SMG, submandibular gland; wCV, within-subject coefficient of variation.

Tissue	RC (95% CI) [ml/100 g/min]	wCV [-]
Parotid Gland Right	23.9 (20.2, 29.0)	16.1%
Parotid Gland Left	21.3 (18.0, 25.9)	13.3%
SMG Right	26.7 (22.5, 32.4)	24.6%
SMG Left	26.1 (22.0, 31.6)	21.8%
Tonsil Right	18.8 (15.4, 23.5)	22.8%
Tonsil Left	18.8 (15.4, 23.8)	20.0%
Cerebellum	10.7 (9.0, 13.0)	9.0%

The average T1 (\pm SD) derived and used as fixed T1 was 1255 \pm 218 ms for the parotid glands, 1584 \pm 188 ms for the submandibular glands, 2317 \pm 191 ms for the tonsils, and 1506 \pm 101 ms for the cerebellum. While the median difference between using a T1 map, compared with a fixed T1, was close to zero in all cases, for individual acquisitions the difference in BF ranged from -34 to 23 ml/100 g/min for the parotid glands, from -23 to 8 ml/100 g/min for the submandibular glands, from -14 to 5 ml/100 g/min for the tonsils, and from -9 to 8 ml/100 g/min for the cerebellum. This is also shown in Figure 3.S2. The repeatability of the BF measurements based on the fixed T1 was similar to the repeatability of BF estimated using a T1 map. Supporting information Table 3.S1 shows the wCV and RC for the fixed T1 approach.

For the parotid glands notable underestimation of the BF (i.e., zero difference is on the edge or outside the 25%-75% percentile range) was present for all three PLDs when comparing non-transit time corrected BF with transit time corrected BF. Notable underestimation of BF was present in the submandibular glands and the left tonsil for the longest PLD (2479 ms) only. For the cerebellum, underestimation was present for the two shortest PLDs (1000 ms and 1632 ms), but not for the longest PLD. The errors in BF when no transit time correction was applied are also shown in Figure 3.S3.

3.4 Discussion

ASL is a perfusion MRI technique of interest for the prediction and monitoring of tumor and healthy tissue response to head and neck (chemo)radiotherapy, with the advantage, especially in the case of repeated examinations, that no external contrast agent is necessary. However, literature on ASL in the head and neck region remains scarce. To gain insight into the applicability of ASL in

the head and neck area, we derived the nominal values and repeatability of the BF obtained from multi-delay pCASL in a variety of tissues in 20 healthy volunteers. We found the average BF in the parotid glands to be the highest (range 33.9–90.0 ml/100 g/min) out of the assessed tissues in the head and neck, with the best wCV (13.3%–16.6%). The submandibular glands (range 20.6–80.6 ml/100 g/min) and the tonsils (range 14.8–45.8 ml/100 g/min) had a lower BF. The wCV of the tonsils and submandibular glands were similar, ranging from 20.0% to 24.6%.

The ROI in the cerebellum showed values of BF and T1 that agreed with previously published literature [104, 105]. When comparing the repeatability in terms of wCV between the cerebellum and the other tissues, the cerebellum had notably better repeatability. One explanation could be the presence of motion in the neck region caused by breathing, swallowing, or jaw movement, which would have affected the tonsils, submandibular glands and parotid glands, but not the cerebellum. The development of more motion-robust ASL sequences, or motion correction for this type of ASL sequence, could therefore be of interest for future research. Another explanation for the varying repeatability across tissues could be the size of the ROIs, where the influence of noise on the ROI average is less for the large cerebellum ROI than for the small tonsil ROI.

The estimated repeatability coefficients were in the range of 18–27 ml/100 g/min depending on the tissue, so if a difference smaller than the RC is detected, it might not be a true difference. Although the RC was relatively high (in the range of 15%–20% of the mean) compared with the absolute values of BF, differences between volunteers in the order of the estimated RC were observed. Additionally, the differences in BF between benign and malignant, or residual tumor and local control cases reported in previous publications, are in the range of the RC found in this study or larger. For example, Abdel Razek [106] reported an average difference between benign and malignant parotid lesions of 33.4 ml/100 g /min which is similar to the RC we estimated for the parotid glands. Fujima et al. [51] reported an average difference of 28.5 ml/100 g/min when comparing residual tumor with local control after treatment of head and neck squamous cell carcinoma. Furthermore, they also reported an average reduction (96.5 ml/100 g/min) in BF from an average tumor BF of 121.4 ± 27.8 ml/100 g/min pretreatment to an average tumor BF of 24.9 ± 14.9 ml/100g/min post-treatment. Because this change in BF is substantially larger than the observed RC, this suggests that the repeatability is sufficient to assess BF changes in a treatment setting [103]. However, in the current study, we obtained intra-session measurements of BF, and therefore we could not assess the stability of BF over the course of a few weeks, a similar time frame to treatment.

The range in average BF measured over the volunteers was approximately the same for the left and right side for the different organs. However, the left tonsil showed a slightly larger BF on average. This is most likely an effect of the three volunteers who displayed a higher BF (≥ 10 ml/100 g/min) in the left tonsil compared with the right tonsil. In a substantial part of the volunteers (10-20%), the range of BF values found did not overlap between the left and right for each of the three organs in the head and neck region. However, in these cases, the absolute differences were smaller than the estimated RC, and therefore it cannot be concluded that this is a true difference. While eroding the delineation of the tissues most likely diminished the influence of delineation and matching uncertainty, subtle differences in the ROIs could still lead to substantial differences in BF measurements, especially in the presence of vessels, such as in the parotid glands. Additionally, the subjects were carefully positioned as straight as possible in the scanner. However, small differences between left and right might still arise from positioning variation.

The T1 value of the tissue of interest highly affected BF quantification. As we show in this study, T1 can vary substantially between tissues, and therefore it is important to use a suitable T1 for each of the tissues. Using a T1 map for this purpose yields a personalized voxel-wise T1, but T1 maps can introduce additional noise in the BF measurements. Using a fixed cohort-average value of T1 for each of the tissues did not yield substantial average bias in BF compared with using a T1 map. However, substantial individual differences were still observed. It is unclear if the benefit of reduced bias (compared with a fixed T1) when using a T1 map outweighs the increased variance of BF measurements because of the introduction of extra noise through the T1 map.

The average transit times we found for each of the tissues in the head and neck region were similar, which suggests that it is sufficient to assess the BF in these tissues simultaneously with only one PLD without introducing substantial bias to the BF measurements, in which case a PLD between 1000 and 1632 ms would be the most appropriate. Assessing these tissues with one PLD provides an opportunity to acquire more averages within the same scan time, and therefore to yield a more stable BF measurement. Still, individual measurements can deviate substantially, especially when an ROI contains multiple tissue types. Although, from these results, it is unclear whether this is caused by uncertainties in the transit time measurements, or because of variations in SNR or effective label duration, or is attributable to true variation in BF. Additionally, to address other areas like base of skull malignancies, which are closer to the cerebellum and for which the transit times are longer, another PLD would be required. It should be noted, however, that the positioning of the labeling plane plays an important role in the expected transit time.

Given the proximity to the feeding arteries, it is necessary to choose a short PLD. We chose a minimum PLD of 1000 ms, similar to Lin et al. [107] and lower than most other reported PLDs for the head and neck region [86]. However, because of shorter distances between the arteries and labeling plane, and the target tissues, the optimal PLD may vary [86]. So a limitation of this study is that no PLDs below 1000 ms were used, because of the large vascular artifacts we observed at lower PLDs. A minimum PLD of 1000 ms could not fully avoid vascular artifacts because of the many large vessels supplying the head, neck, and brain. Therefore, vascular artifacts could have potentially influenced the BF quantification. Additionally, no upper limit for the BF was used in the analyses, therefore, it is possible that intra-arterial signal could have influenced the BF quantification. In the future, taking account of the arterial component by using a two-compartment model could present a solution to this problem [108].

To exclude the non-physiological result of a negative BF, voxels containing negative BF were excluded from the analysis. As demonstrated in Tables 3.S2 and 3.S3, including voxels with a negative BF would have led to a deterioration of repeatability, while the nominal BF values were hardly affected (maximum deviation <2.2 ml/100 g/min, which was well within one SE).

In the current analysis, full perfusion maps were constructed and, subsequently, ROI averages were computed. Because of the non-linear model fitting, this approach is probably more susceptible to acquisition noise than performing the model fitting on ROI-averaged signals. However, that would assume homogeneous tissues, which we did not want to assume because of the intended application in tumors, which are, in general, heterogeneous structures.

In this work, a blood-tissue partition coefficient (λ) of 0.9 was used, as is customarily used for brain imaging. However, a value of 1 was also previously used in studies in head and neck cancer [51, 109]. Because the estimated BF scales linearly with λ , an error in λ of 10% yields an error of 10% in BF as well. However, it could be that the tissues assessed in this study have different values for λ , as λ is known to vary per age, tissue, subject, and pathology [110, 111].

Finally, the volunteers in this cohort were all healthy, young volunteers. BF can be influenced by several diseases and may deviate by age. Therefore, caution should be taken when translating these values to other cohorts. Additionally, expanding the cohort would have likely resulted in a smaller SE of the mean. However, the gain in SE would most likely be limited because currently, it is around 5% of the mean.

3.5 Conclusion

We demonstrated the feasibility of using multi-delay pCASL to simultaneously measure BF in several tissues in the head and neck region. Moreover, we presented nominal BF values for the parotid glands, submandibular glands, and tonsils. While some variation in BF was found between different subjects, generally the parotid glands were the most perfused and the tonsils the least perfused. Additionally, we have also reported the average values of T1 needed for BF quantification for the tissues assessed, which, to the best of our knowledge, have not been reported before. The results suggest that the repeatability of the BF measurements we found is sufficient for these measurements to be included in clinical studies regarding tumor detection, and treatment and toxicity monitoring.

3.6 Supporting information

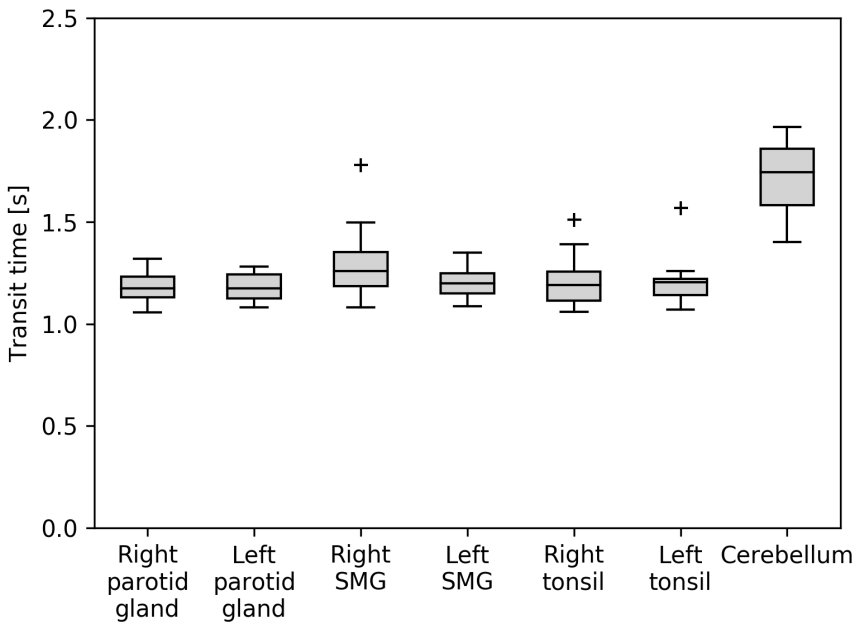


Figure 3.S1. Boxplot of the mean transit time per volunteer (mean over the repeats) for each of the tissues of interest. All tissues show similar transit times with exception of the cerebellum. SMG – submandibular gland.

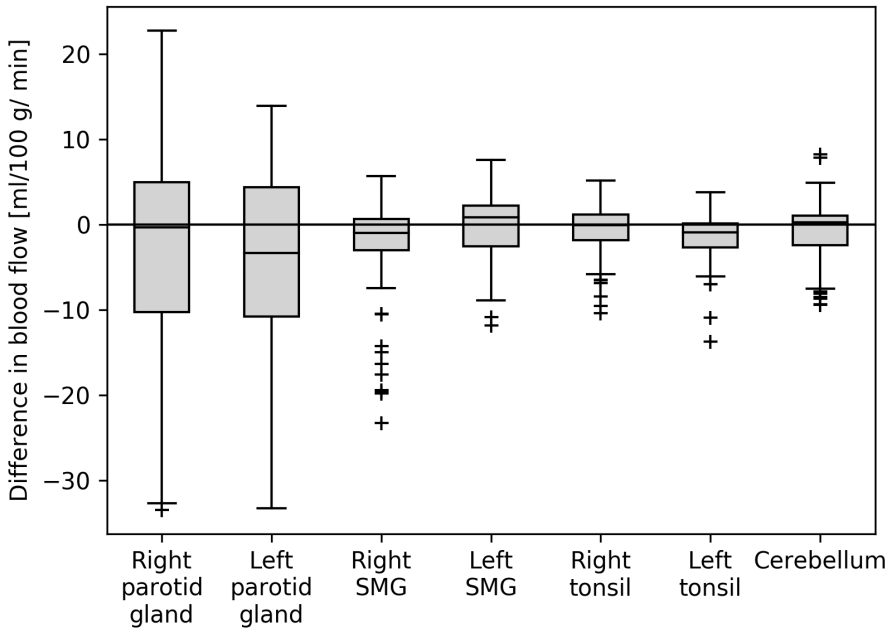


Figure 3.S2. Boxplot of the difference between using a fixed T1 and a T1 map for BF estimation in the tissues of interest. Notably, the median differences tend to be close to zero, but in the parotid glands a larger spread is found compared to the other tissues. SMG – submandibular gland.

Table 3.S1. Repeatability measures and confidence interval of BF for the assessed tissues. CI – confidence interval, RC - repeatability coefficient, SMG - submandibular gland, wCV – within-subject coefficient of variation.

Tissue	T1 map		Constant T1	
	RC (95% CI) [ml/100 g/min]	wCV [-]	RC (95% CI) [ml/100 g/min]	wCV [-]
Parotid Gland Right	23.9 (20.2, 29.0)	16.1%	21.4 (18.0, 25.9)	15.0%
Parotid Gland Left	21.3 (18.0, 25.9)	13.3%	20.0 (16.9, 24.2)	13.3%
SMG Right	26.7 (22.5, 32.4)	24.6%	25.2 (21.3, 30.5)	24.7%
SMG Left	26.1 (22.0, 31.6)	21.8%	24.9 (21.0, 30.2)	20.7%
Tonsil Right	18.7 (15.4, 23.5)	22.8%	17.5 (14.4, 22.0)	21.8%
Tonsil Left	18.8 (15.4, 23.8)	20.0%	18.6 (15.2, 23.5)	20.8%
Cerebellum	10.7 (9.0, 13.0)	9.0%	10.6 (9.0, 12.9)	9.1%

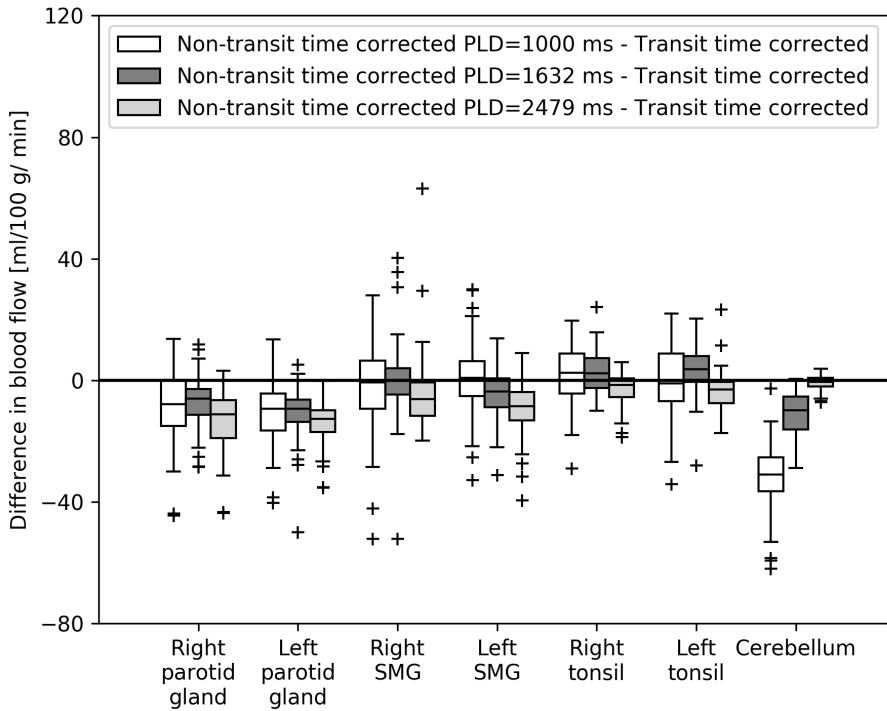


Figure 3.S3. Boxplot of the difference between non-transit time corrected BF at each of the PLDs and the transit time corrected BF. Note that for the submandibular gland and left tonsil for the longest PLD the median falls just outside or on the edge of the box of the boxplot (25%-75% percentiles) and underestimation of the blood flow is generally found. This is also still the case for all PLDs in the parotid glands. Underestimation was also found for the first two PLDs for the cerebellum. SMG – submandibular gland.

Table 3.S2. Nominal values of the BF with the corresponding SD and standard error (SE) with negative voxels included in the analysis.

Tissue	BF [ml/100g/min]	SD	SE
Parotid glands (n=20)	54.9	14.4	3.2
Submandibular glands (n=20)	39.0	13.5	3.0
Tonsils (n=14)	32.2	8.1	2.2
Cerebellum (n=20)	42.8	10.8	2.4

Table 3.S3. Repeatability measures and confidence interval of BF for the assessed tissues with negative voxels included in the analysis. CI – confidence interval, RC – repeatability coefficient, SMG – submandibular gland, wCV – within-subject coefficient of variation.

Tissue	RC (95% CI) [ml/100 g/min]	wCV [-]
Parotid Gland Right	25.5 (21.5, 30.8)	17.6%
Parotid Gland Left	23.4 (19.8, 28.4)	14.7%
SMG Right	32.6 (27.5, 39.5)	32.7%
SMG Left	28.1 (23.7, 34.1)	24.1%
Tonsil Right	18.8 (15.5, 23.5)	22.8%
Tonsil Left	19.1 (15.6, 24.1)	20.3%
Cerebellum	11.9 (10.1, 14.4)	10.0%

5

The COMPLETE trial: Holistic early response assessment for oropharyngeal cancer patients; Protocol for an observational study

Gerda M. Verduijn, Marta E. Capala,
Nienke D. Sijtsema, Iris Lauwers,
Juan A. Hernandez Tamames,
Wilma D. Heemsbergen, Aniel Sewnaik,
Jose A. Hardillo, Hetty Mast, Yvette van Norden,
Maurice P.H.M. Jansen, Aad van der Lugt,
Dik C. van Gent, Mischa S. Hoogeman,
Bianca Mostert, Steven F. Petit

This chapter is based on Verduijn GM, Capala ME, Sijtsema ND, Lauwers I, Hernandez Tamames JA, Heemsbergen WD, et al. The COMPLETE trial: Holistic early response assessment for oropharyngeal cancer patients; Protocol for an observational study. BMJ Open. 2022;12:e059345

Abstract

Introduction

The locoregional failure (LRF) rate in human papillomavirus (HPV)-negative oropharyngeal squamous cell carcinoma (OPSCC) remains disappointingly high and toxicity is substantial. Response prediction prior to or early during treatment would provide opportunities for personalized treatment. Currently, there are no accurate predictive models available for correct OPSCC patient selection. Apparently, the pivotal driving forces that determine how a OPSCC responds to treatment, have yet to be elucidated. Therefore, the holistic early response assessment for oropharyngeal cancer patients study focuses on a holistic approach to gain insight into novel potential prognostic biomarkers, acquired before and early during treatment, to predict response to treatment in HPV-negative patients with OPSCC.

Methods and analysis

This single-center prospective observational study investigates 60 HPV-negative patients with OPSCC scheduled for primary radiotherapy (RT) with cisplatin or cetuximab, according to current clinical practice. A holistic approach will be used that aims to map the macroscopic (with Intra Voxel Incoherent Motion Diffusion Kurtosis Imaging (IVIM-DKI); before, during, and 3 months after RT), microscopic (with biopsies of the primary tumor acquired before treatment and irradiated ex vivo to assess radiosensitivity), and molecular landscape (with circulating tumor DNA (ctDNA) analyzed before, during and 3 months after treatment). The main endpoint is locoregional control (LRC) 2 years after treatment. The primary objective is to determine whether a relative change in the mean of the diffusion coefficient D (an IVIM-DKI parameter) in the primary tumor early during treatment, improves the performance of a predictive model consisting of tumor volume only, for 2-years LRC after treatment. The secondary objectives investigate the potential of other IVIM-DKI parameters, ex vivo sensitivity characteristics, ctDNA, and combinations thereof as potential novel prognostic markers.

Ethics and dissemination

The study was approved by the Medical Ethical Committee of Erasmus Medical Center. The main results of the trial will be presented in international meetings and medical journals.

Trial registration number

NL8458.

5.1 Introduction

Head and neck cancer (HNC) is the sixth most common type of cancer worldwide with an estimated annual burden of 633 000 new cases and 355 000 deaths [125]. Despite recent advances in treatments resulting in better outcomes for diseases such as melanoma or lung cancer, the treatment of HNC continues to disappoint, especially for human papillomavirus (HPV)-negative HNC. Blanchard et al. reported a 2-year overall survival (OS) of 50.7% for the chemoradiotherapy group, and 46.0% after radiotherapy (RT) alone in their meta-analyses on HPV-negative oropharyngeal squamous cell carcinoma (OPSCC) [126]. Ang et al. reported a 3-year locoregional recurrence rate of 35.1% in the HPV-negative OPSCC group [127]. This rate indicates that a considerable number of patients die due to locoregional recurrence for which there are no other curative treatment options in the majority of cases. Furthermore, the burden of acute and late side effects is still substantial despite the introduction of modern radiation techniques [128-131].

Currently, 650 new patients with OPSCC are diagnosed annually in the Netherlands of which 40%–50% are HPV-negative. If we could predict treatment response in this patient group before or early during treatment, this would open the door to clinical trials in which a more personalized treatment could be investigated, for example, intensified (or in contrast, for those with poorer performance status, palliative therapy) for poor responders, and possibly less intense and thereby a less toxic therapy for good responders. Although there have been studies performed to determine prognostic factors for patients with HNC [132-137], to date no accurate predictive model exists for patients with HPV-negative OPSCC for a number of reasons. (1) Previous studies have focused mainly on patient/clinical characteristics (tumor volume, age, smoking history, comorbidities) in addition to biomarkers of a maximum of one modality (e.g., MRI), while the response of the tumor depends on its entire, complex, multilayered landscape [138]. (2) Many studies focused on pretreatment characteristics only, while a tumor is a dynamic system that changes during treatment. (3) Studies are too small ($n \sim 30$) and contain patients with different types of head and neck tumors as well as HPV-negative and HPV-positive tumors combined.

The current holistiC early respOnse assessMent for oroPharyngeaL cancer paTiEnts (COMPLETE) study was designed to address these shortcomings directly by (1) studying the entire multilayered tumor landscape based on novel techniques focusing on the macroscopic, microscopic, and molecular landscape; (2) assess changes in the tumor landscape early during treatment

and (3) acquire data in a cohort consisting of 60 patients with HPV-negative OPSCC, respectively.

The *macroscopic* tumor landscape will be studied with multi-b-value diffusion-weighted imaging (DWI) using the hybrid Intra Voxel Incoherent Motion Diffusion Kurtosis Imaging (IVIM-DKI) model [42, 124]. With DWI the extracellular movement of water molecules is detected and quantified by the apparent diffusion coefficient. When adding the IVIM-DKI model, perfusion and intracellular diffusion (reflected by the kurtosis) are taken into account. Obtaining additional parameters from DWI by employing IVIM and DKI will enlarge the potential of macroscopic response prediction. This multi-b-value DWI sequence will be obtained before, during, and after treatment to study changes over time [80, 121].

For the *microscopic* landscape, ex vivo radiosensitivity assessment of patient-specific tumor biopsies will be obtained before treatment as a potential biomarker of clinical outcome. We recently adapted our breast cancer organotypic tumor tissue slice method to be suitable for head and neck tumor tissue (publication in preparation) and developed a protocol for ex vivo radiation treatment of tumor tissue [139]. Using this method, tumor sensitivity to irradiation can be assessed for each individual patient.

Finally, the *molecular* landscape will be studied by analyzing liquid biopsies and collecting circulating tumor DNA (ctDNA) for molecular tumor characteristics before, during, and after treatment. Liquid biopsies are a promising minimally invasive alternative for tissue biopsies and serial samples at different time points during treatment are easily acquired. ctDNA comprises DNA fragments derived from tumor cells, which enter the bloodstream after apoptosis or by active shedding of DNA fragments by living tumor cells. Genetic aberrations, such as mutations, can be identified and tracked in ctDNA, and correlated with clinical outcomes. In several tumor types, ctDNA detected at baseline and its evolution during treatment were shown to be strong prognostic factors [140–142]. Wang et al. were able to detect ctDNA in the plasma of HNC patients in a proof-of-principle study. In a small subgroup that did not develop tumor recurrence, no mutations were present shortly after primary surgery [143]. This makes the detection of ctDNA a potential early biomarker that can be used to further tailor treatment.

5.2 Methods and analysis

5.2.1 Design and study population

The COMPLETE study is a single-center prospective observational study. In the period of August 2020 until August 2024, 60 patients will be included with histologically proven cT1-2N2-3M0 or cT3-4N0-3M0 HPV-negative OPSCC treated with primary RT and chemotherapy (cisplatin) or epidermal growth factor receptor (EGFR)-targeted therapy (cetuximab). For the choice of number of patients, we refer to the power calculation in the statistical section.

5.2.2 Study objectives

Primary objective

Among the biomarker modalities explored in the current study (DWI, ex vivo radiosensitivity, and ctDNA), most data are available on DWI parameters in relation to treatment outcome. Therefore, the primary objective of the study will be to determine if a relative change in the mean of the diffusion coefficient D (as obtained from IVIM-DKI) in the primary tumor early during treatment improves the performance of a predictive model consisting of only tumor volume for the 2-years locoregional control (LRC) after treatment of patients with HPV-negative OPSCC.

Secondary objectives

1. To determine if a relative change in the mean of the diffusion coefficient D in the primary tumor early during treatment improves the performance of a predictive model including tumor volume only for the 3 months response after treatment of patients with HPV-negative OPSCC.
2. To determine if other IVIM-DKI parameters (perfusion fraction f , pseudo-diffusion coefficient D^* , and kurtosis K), ctDNA, ex vivo radiosensitivity characteristics, and combinations thereof can be identified as potential novel predictive markers for treatment response of patients with HPV-negative OPSCC, using an explorative approach.
3. To build a repository of imaging data and liquid biopsies to allow future identifications of biomarkers of treatment response of patients with HPV-negative OPSCC.

5.2.3 Inclusion criteria

- Patients with histologically proven cT1-2N1-3M0 or cT3-4N0-3M0 HPV-negative OPSCC.

- Eighteen years or older.
- Current and/or former smoker.
- Scheduled for primary RT with chemotherapy (cisplatin) or EGFR-targeted therapy (cetuximab).
- Standard planning MRI (including IVIM-DKI) successfully acquired.
- Included in the BIOmarker of treatment Response in Oropharyngeal Cancer (BIO-ROC) study (see supporting information)
- Written informed consent.

5.2.4 Exclusion criteria

- Patients with recurrence of previously confirmed head and neck squamous cell carcinoma or with other malignancies within the last 5 years.
- Patients with previous irradiation or surgery in a head and neck region overlapping with the current tumor.
- Patients with any physical or mental status that interferes with the informed consent procedure or study procedures.
- Patients with contraindications for MRI (e.g., claustrophobia, arterial clips in the central nervous system).
- Patients with contraindications for gadolinium contrast (i.e., hypersensitivity for gadolinium or impaired kidney function).
- We will continue inclusion until we have 60 evaluable subjects, that is, with the required MRI scans and blood samples.

5.2.5 Study procedures

The general outline of the study procedures is presented in Figure 5.1. Patients will be discussed in the weekly meeting of the multidisciplinary head and neck tumor board and patients will be treated according to the current clinical protocols. Patients will receive 70 Gy intensity modulated radiotherapy or intensity modulated proton beam therapy in 35 fractions combined with cisplatin (100 mg/m² body surface area (BSA), once every 3 weeks or 40 mg/m² BSA, every once a week)

or cetuximab (initial dose of 400 mg/m², followed by 250 mg/m² weekly, for the duration of RT).

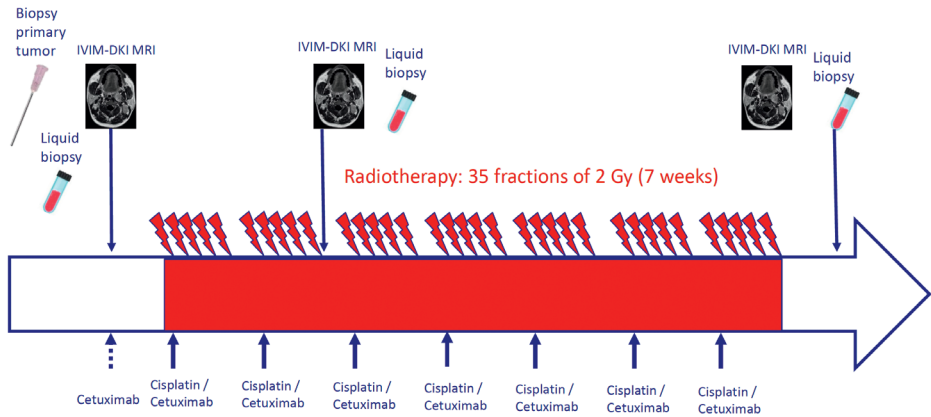


Figure 5.1. Standard clinical procedures for patients with oropharyngeal squamous cell carcinoma treated with primary radiotherapy with cisplating or cetuximab in our center, as well as the study procedures of the *holistiC* early respOnse assessMent for oroPharyngeAL cancer paTiEnts trial. The procedures that are specific for the study are an additional tumor biopsy and a liquid biopsy (circulating tumor DNA (ctDNA)) before treatment. The MR scanning session, including an Intra Voxel Incoherent Motion Diffusion Kurtosis Imaging (IVIM-DKI) diffusion-weighted MRI sequence, that is part of the clinical protocol is repeated as part of the study in the second week of treatment, and 3 months after radiotherapy. At the same time points, a second and third liquid biopsy (ctDNA) is acquired.

Timing of study procedures

Eligible patients are asked to participate in the BIO-ROC study (see supporting information). As part of the BIO-ROC study, a study-specific biopsy, and a blood sample of 30 mL will be obtained before the start of treatment. An MRI scan will be performed before the start of treatment as part of the standard workup. In the second week of treatment, a blood sample will be acquired for ctDNA analysis and the patient will undergo a second MRI scan. Three months after the completion of RT, at the time of clinical response evaluation, a third blood sample will be acquired for ctDNA analysis and the patient will undergo a third MRI scan.

The macroscopic landscape: IVIM-DKI

MRI scans will be acquired with the patient immobilized in treatment position (i.e., with an RT mask). The MRI scan protocol consists of T1-weighted (T1w) Dixon after gadolinium contrast material injection, a T2-weighted (T2w) turbo spin echo, a multi-b-value DWI scan and a DWI scan with inverse phase

encoding gradient polarity for the purpose of distortion correction (flip angle: 90 degrees; repetition time (TR): 6700 ms; echo time (TE) 81.8 ms; field of view (FOV) 26×26 cm; 4 mm slice thickness; 0.2 mm gap, 128×128 matrix; bandwidth: 1953.12 Hz/pixel). The multi-b-value DWI scan consists of 15 b-values (0, 10, 2×80, 130, 570, 2×770, 2×780, 790, and 4×1500 s/mm²) acquired in three orthogonal diffusion directions [121], where the b-values represent the amount of diffusion weighting.

The microscopic landscape: biopsy

For patients with a tumor that is accessible during physical examination (with or without histological confirmation), a tumor biopsy will be obtained by a head and neck surgeon during the outpatient clinic visit according to the BIO-ROC study (see supporting information). For patients without histology-confirmed OPSCC, and requiring general anesthesia for proper tumor approach, two biopsies will be obtained during a single procedure, one for the diagnosis and one for the purpose of the study. The tumor biopsies will be sliced into 300 μM thick slices, irradiated ex vivo, and cultured for 5 days. Based on preliminary results from our laboratory, a single dose of 5 Gy resulted in the best discrimination between irradiation-sensitive and irradiation-resistant tumors [144]. Therefore, all tumor biopsies (of individual patients) used in the current study will be treated with a single dose of 5 Gy. In case more tumor material is available allowing for multiple treatment conditions, separate slices of the same tumor will also be treated with a single dose of 2 Gy or 7 Gy to gain more insight into the irradiation sensitivity of a given tumor.

The molecular landscape: ctDNA blood samples

Blood samples containing 30 mL blood for ctDNA analysis will be stored in CellSave tubes for ctDNA analysis at room temperature until processing it to plasma. Subsequently, cell-free DNA (cfDNA) will be isolated using the manual QIAmp circulating nucleic acid kit (Qiagen) or the automated QIASymphony (Qiagen) or Maxwell kits (Promega). The plasma and isolated cfDNA will be stored at -80 °C and -30 °C, respectively, until further analysis.

5.2.6 Patient follow-up

Patients are monitored by the head and neck multidisciplinary team according to national guidelines. Follow-up visits will be planned every 2 months for the first year following RT. Starting from the second year, the frequency gradually decreases to every 6 months for a minimum of 5 years. LRC at 2 years will be determined by clinical examination and in case of doubt additional imaging and/or biopsies will be acquired according to current clinical practice.

5.2.7 Data processing and analysis

The macroscopic layer: IVIM-DKI analysis

The primary tumor will be delineated on the pretreatment T1w and T2w scans. The multi-b-value DWI acquisitions will be processed according to Sijtsema et al. [121]. In short, first, the scans for each b-value will be corrected for geometric distortion with FSL (FMRIB Software Library) [47, 76]. Second, the scans of the individual b-values are registered rigidly to the scan with $b=0$ s/mm². Note that a rigid registration is expected to suffice since patients are scanned with the RT mask. Then the region of interest (ROI), as defined by the primary tumor contours, is projected on top of the scan with $b=0$ s/mm². Then the diffusion coefficient values are calculated for each voxel in the ROI by fitting the IVIM-DKI model based on different b-values from the multi-b-value DWI acquisition:

$$S_i = S_0 \left((1 - f) e^{-b_i D + \frac{1}{6} (b_i D)^2 K} + f e^{-b_i D^*} \right) \quad (5.1)$$

where S_i is the measured signal intensity at the corresponding b-value b_i , S_0 is the signal intensity at a b-value of 0 s/mm², D is the diffusion coefficient, f is the perfusion fraction, D^* is the pseudo-diffusion coefficient, and K is the kurtosis. The b-values represent the amount of diffusion weighting. The mean diffusion coefficient D of the ROIs will be calculated for both the pretreatment scans (acquired as part of the clinical protocol) and the scans acquired in the second week of treatment. The percentage change in mean diffusion coefficient D during treatment compared with pretreatment is used for the statistical analysis of the primary endpoint. Next, for D , f , D^* , and K the distribution within the tumor is calculated. From the distribution, a large variety of metrics will be extracted, among others the SD, and the 80th, 90th, 95th, and 99th percentiles, which will be used as input for an exploratory analysis. Moreover, supervoxels will be created to analyze the heterogeneity in the tumor.

The microscopic layer: ex vivo radiation and radiosensitivity testing

The percentage of proliferating cells of the irradiated tumor slices will be compared with untreated tumor slices after 5 days of culture. Proliferation will be detected by 5-ethynyl-2-deoxyuridine (EdU) incorporation and obtained microscopy images will be analyzed using in-house image processing software (Apoptosis Quantifier) for semi-automated quantification of the results. Similarly, an increase in apoptosis in irradiated slices will be assessed after 5 days, using terminal deoxynucleotidyl transferase dUTP nick-end labeling (TUNEL). Untreated slices will be used as a control. The same in-house processing software will be used for microscopy image analysis. The outcomes

of both assays will be analyzed as a continuous variable in the exploratory statistical analysis. Change in both parameters compared with the control will be used to describe tumor irradiation sensitivity.

The molecular layer: ctDNA analysis

A targeted approach with molecular barcoding will be applied using a panel of somatic genetic variations, including TP53, PIK3CA, CDKN2A, FBXW7, HRAS, NRAS, FAT1, and MOTCH1 [143, 145]. This panel will be extended based on the most recent available primary tumor sequencing data and literature at the time of analysis, which will be expected to cover the relevant genetic aberrations of interest in HPV-negative OPSCC.

At least 20 ng of cfDNA will be sequenced using the above mentioned customized panel with molecular barcoding on the Ion Torrent NGS platform. The molecular barcoding will enable molecule quantification and detect mutations as low as 0.1% allele mutation frequency when evaluating 20 ng of cfDNA input. The TorrentSuite variant calling pipeline is used to identify tumor-specific variants for ctDNA detection, including TP53 variants, and quantify the number of reads and independent molecules with wild-type and variant sequences. Subsequently, based on these reads and molecule levels, the variant allele frequency and the number of mutant molecules per mL blood will be established. DNA from the buffy coat will also be isolated and sequenced with this panel, to identify germline variants and mutations due to clonal hematopoiesis.

The ctDNA extraction and analysis will be performed on the blood samples acquired pretreatment, acquired in the second week of treatment, and acquired at 3 months post-treatment. The change in the total number of mutant molecules in week 2 compared with baseline, specific genetic variants, the total number of mutations, the total ctDNA concentration in the blood, and how these evolve during treatment will be described.

5.2.8 Statistical analyses

Primary objective

The dependent variable is LRC at 2 years (yes/no). Based on relevant literature [134], within our study population of patients with HPV-negative OPSCC and a smoking history, 37% of the patients are expected to have local tumor progression within 2 years (the primary outcome of interest). We expect to be able to include 60 patients in 4 years, which will lead to approximately 22 events in total. Twenty-two events allow the testing of two explanatory variables based on the rule of thumb that 10 events are required per variable. In case of

missing values, the analyses will be done on the complete cases for the specific analysis but with sensitivity analyses after imputation on all included patients.

A multivariable logistic regression will be performed with LRC at 2 years as dependent variable. According to literature, tumor volume based on the delineated gross tumor volume pre-RT is the most important variable associated with LRC 2 years after treatment among our patient population of only HPV-negative OPSCC patients treated with primary RT with chemotherapy (cisplatin) or EGFR-targeted therapy (cetuximab) [132, 133, 146-148]. The second variable that will be included is the relative change in mean diffusion coefficient D in week 2 compared with baseline as determined by the IVIM-DKI scans. The multivariable model including both parameters will be compared with the model without the change in mean diffusion coefficient D . A likelihood ratio test will be applied to determine if the model with the change in mean diffusion coefficient D performs better than the model without; where a p -value <0.05 will be considered statistically significant.

Secondary objectives

The first secondary objective is, apart from the endpoint at 3 months instead of 2 years, equivalent to the primary objective; the statistical analysis is therefore identical to the one described for the primary endpoint. The analysis for the first secondary objective will be performed once the 3-month endpoint is reached for all patients.

For the other secondary objectives, the parameters that will be analyzed include:

- Clinical/patient characteristics such as age, comorbidities, and clinical tumor stage.
- IVIM-DKI parameters D , f , D^* , K and their distributions within the tumor (at baseline and in week 2). Moreover, supervoxels will be generated based on the combination of D , f , K , and D^* to investigate the effect of different distinct tumor regions on LRC.
- The established ex vivo radiosensitivity parameters (changes in proliferation and apoptosis on irradiation with different irradiation doses).
- ctDNA parameters such as the total number of mutant molecules, the presence of specific genetic variants, the total ctDNA concentration in the blood, and how these evolve during treatment.

Different endpoints will be considered: LRC at 3 months, LRC at 2 years, and OS at 2 years.

Given the large number of variables compared with the number of events, feature selection is necessary but the risk of overfitting is significant. As conventional statistics are not suitable for the secondary objectives, an exploratory analysis will be performed using Least Absolute Shrinkage Selector Operator (LASSO) logistic regression. LASSO logistic regression is a type of regression that shrinks the coefficients of the variables to avoid overfitting, while performing feature selection at the same time. Furthermore, LASSO is a good balance between conventional statistical approaches, such as backward selection, and more black-box, data-driven machine learning techniques. Analysis will be performed with the penalized package in R Statistical software. We will use L1 regularization given the large number of variables tested. Internal validation will be performed with cross-validation. In correspondence to the primary hypothesis, in case of missing values, the analyses will be done on the complete cases for the specific analysis but with sensitivity analyses after imputation on all included patients.

5.2.9 Patient and public involvement

The Dutch patient association for head and neck cancer (PVHH) gave feedback on our project during the development phase and will continue to provide feedback during the trial.

5.2.10 Ethics and dissemination

The study was approved by the Medical Ethical Committee of Erasmus Medical Center (MEC 2020-0208). The COMPLETE trial is supported by the Dutch patient association for head and neck cancer (PVHH). The methods and findings of the study will be published in peer-reviewed journals and presented at national and international conferences.

5.3 Discussion

Although several strategies implemented in recent years in the treatment of patients with OPSCC have increased LRC, there is still an urgent need for improvement, especially for patients with HPV-negative OPSCC. To be able to select the right patient for treatment intensification or de-intensification, an accurate predictive model needs to be developed. Given the complexity and the dynamics of tumor response as an interaction between the different 'layers' (macroscopic, microscopic, and molecular) that evolve as a result of treatment, we believe that for accurate prediction models the different layers and the dynamics of response should be incorporated. In the current COMPLETE

study, we aim to assess the entire multilayered tumor landscape based on novel techniques focusing on the macroscopic, microscopic, and molecular landscape before and early during treatment, in a patient cohort containing 60 patients with HPV-negative OPSCC only.

There is a delicate balance between acquiring as much information as possible before and during treatment, while limiting the number of procedures patients need to undergo. For the macroscopic data, we chose to focus on the novel IVIM-DKI MRI technique, since conventional DWI has shown to be promising for response assessment of HNC [36, 149-151]. IVIM-DKI adds information compared with conventional DWI but also has limitations. For instance, Sijtsma et al. demonstrated a relative repeatability coefficient of the diffusion coefficient D of 38% in healthy volunteers [121]. So, fairly large changes in D need to occur to be detected as a true change, as small changes will be within normal measurement variation. As an alternative, several other functional imaging modalities could have been candidates to provide early response assessment as well for the macroscopic layer, for example, fluorodeoxyglucose positron emission tomography (PET)-CT [152]. Our decision to focus on MRI was based on prior studies [36, 149], that MRI is part of our standard workflow in RT planning for HNC, and therefore does not require an additional scanning session pretreatment and the short scanning time resulting in manageable patient discomfort. Possibly, adding one or two PET-CT on top of the MRI scans would have provided additional interesting data, but was deemed infeasible regarding the additional patient burden.

For microscopic data, we study the response of tumor biopsies to irradiation *ex vivo*. This novel technique might have profound clinical implications, allowing individualized treatment of patients with OPSCC. However, for several reasons, *ex vivo* response may not turn out to be representative for patient response. For instance, the biopsy may not represent intra-tumor heterogeneity of a tumor that may consist of different regions. Furthermore, tumor tissue is grossly selected at the outpatient clinic without microscopic confirmation potentially yielding tissue with low cellularity. However, based on our experience so far, the risk of missampling is small.

For the molecular data, we focus on ctDNA as this is a promising biomarker that is easily acquired [140-143]. A possible limitation of ctDNA is the detection of DNA fragments at very low concentrations. Other possible candidates to assess the molecular landscape would have been circulating tumor cells (CTCs), microRNA (miRNA), and cfRNA. However, since CTCs have so far not been established as a prognostic marker in locally advanced HNC and the low sensitivity in the primary (non-metastasized) setting, no CTC analyses are part

of the study [153]. miRNAs are also a promising prognostic markers, but are not an area of expertise in our laboratory and were therefore not chosen as a marker. cfRNA as a biomarker is strongly challenged by the need to process blood samples quickly after a blood draw, which is a challenge logistics-wise.

We expect that, given the complexity of tumor response, the holistic approach we propose is promising to identify combinations of biomarkers for accurate prediction models. Naturally, studying multiple variables has as important drawback the required number of events for sufficient statistical power. Therefore, the study was powered solely on a macroscopic level parameter; the change in mean diffusion coefficient. The secondary objectives that combine multiple parameters from the different layers should therefore be considered as explorative and hypothesis-generating to select a high potential combination of biomarkers to be validated in subsequent trials.

5.4 Supporting information

Ancillary study: The BIO-ROC (BIOmarker of treatment Response in Oropharyngeal Cancer) study

All newly diagnosed OPSCC patients in our medical center are asked to participate in the BIO-ROC study that aims to assess the influence of intrinsic tumor properties on treatment outcomes. This study is a prospective exploratory cohort study for OPSCC patients treated with primary radiotherapy with or without the addition of chemotherapy (cisplatin) or EGFR-targeted therapy (cetuximab) with curative intent. The goal is to assess the correlation between tumor *ex vivo* radiosensitivity with clinical response and to build a database of tumor and blood samples for future biomarker identification. For patients with OPSCC accessible during physical examination an additional tumor biopsy will be obtained pretreatment. For patients without histological confirmation of OPSCC and requiring general anesthesia for tumor approach, an extra biopsy next to the diagnostic one will be obtained during a single procedure. For all patients an additional blood sample will be obtained pretreatment, at the end of week 2 during RT, and three months after RT during the clinical response evaluation visit. Clinical outcomes will be assessed within the standard follow-up scheme. In case of tumor recurrence, patients will be approached for obtaining additional tumor and blood samples. Additional informed consent will be asked for the BIO-ROC patients that meet the inclusion criteria of the COMPLETE protocol.

6

Relating pretreatment non-Gaussian intravoxel incoherent motion diffusion- weighted imaging to HPV status and response in oropharyngeal carcinoma

Nienke D. Sijtsema*, Iris Lauwers*,
Gerda M. Verduijn, Mischa S. Hoogeman,
Dirk H.J. Poot, Juan A. Hernandez-Tamames,
Aad van der Lugt, Marta E. Capala,
Steven F. Petit

*These authors contributed equally

Submitted.

Abstract

Background and purpose

Diffusion-weighted imaging (DWI) is a promising technique for response assessment in head and neck cancer. Recently, we optimized Non-Gaussian Intravoxel Incoherent Motion Imaging (NG-IVIM), an extension of the conventional apparent diffusion coefficient (ADC) model, for the head and neck. In the current study, we describe the first application in a group of patients with human papillomavirus (HPV)-positive and HPV-negative oropharyngeal squamous cell carcinoma. The aim of this study was to relate ADC and NG-IVIM DWI parameters to HPV status and clinical treatment response.

Materials and methods

Thirty-six patients (18 HPV-positive, 18 HPV-negative) were prospectively included. Presence of progressive disease was scored within one year. The mean pretreatment ADC and NG-IVIM parameters in the gross tumor volume were compared between HPV-positive and HPV-negative patients. In HPV-negative patients, ADC and NG-IVIM parameters were compared between patients with and without progressive disease.

Results

ADC, the NG-IVIM diffusion coefficient D , and perfusion fraction f were significantly higher, while pseudo-diffusion coefficient D^* and kurtosis K were significantly lower in the HPV-negative compared to HPV-positive patients. In the HPV-negative group, a significantly lower D was found for patients with progressive disease compared to complete responders. No relation with ADC was observed.

Conclusion

The results of our single-center study suggest that ADC is related to HPV status, but not an independent response predictor. The NG-IVIM parameter D , however, was independently associated with response in the HPV-negative group. Noteworthy in the opposite direction as previously thought based on ADC.

6.1 Introduction

Diffusion-weighted imaging (DWI) is highly interesting for response assessment in head and neck (HN) cancer. Low baseline apparent diffusion coefficient (ADC) has been associated with favorable response to treatment compared to high baseline ADC [37, 154-157]. Non-Gaussian Intravoxel Incoherent Motion Imaging (NG-IVIM) DWI is a novel extension of conventional DWI that enables simultaneous assessment of *inter*-cellular diffusion (similar to the ADC obtained from conventional DWI), microvascular perfusion (like IVIM DWI), and *intra*-cellular diffusion (like diffusion kurtosis imaging) [42]. Compared to conventional DWI, where only the ADC is obtained, NG-IVIM DWI provides a more detailed picture of the tumor micro-environment.

In a previous study [121], we optimized NG-IVIM DWI specifically for the HN region to allow optimal parameter estimation at maximum time efficiency, i.e. with a minimal number of b-values. In the current study, this optimized NG-IVIM DWI sequence was applied for the first time to a group of patients with oropharyngeal squamous cell carcinoma (OPSCC).

The current study includes two important subcategories of OPSCC patients: patients with tumors that are human papillomavirus (HPV)-negative and HPV-positive. On average, HPV-positive patients have a more favorable response to treatment than HPV-negative patients [9]. It may be important to include both categories, since a recent publication [158] suggested that the ADC obtained from conventional DWI might not be an *independent* prognostic factor for response, but rather a surrogate for HPV status. If that would be the case, the value of DWI for response prediction in OPSCC might be lower than expected based on earlier studies [37, 154-157].

The aims of this study were to apply the optimized NG-IVIM DWI sequence for the first time in HPV-positive and HPV-negative OPSCC patients, to study differences in pretreatment conventional DWI and NG-IVIM DWI parameters between HPV-positive and HPV-negative patients, and to relate pretreatment conventional DWI and NG-IVIM DWI parameters to response within one year after treatment.

6.2 Materials and methods

6.2.1 Patients

This prospective study was approved by the institutional review board (protocols 20-0207 and 21-0847) and written informed consent was obtained for all included patients. Patients were eligible if they had OPSCC scheduled

for primary (chemo)radiotherapy; received a radiotherapy planning MRI with a multi b-value NG-IVIM DWI as part of the standard work-up between April 2020 and February 2022; and for which the primary tumor was clearly visible on the DWI image of each b-value. Tumor staging was done according to TNM classification, edition 8.

6.2.2 Treatment

Patients received either volumetric arc photon therapy (VMAT) or intensity-modulated proton therapy (IMPT) of 70 Gy (35 fractions of 2 Gy) to the primary tumor and regions containing pathological neck nodes, and 54.25 Gy to the elective neck regions, with a simultaneous integrated boost. The overall treatment time was either 7 weeks (5 fractions/week) or 6 weeks (6 fractions/week). Chemotherapy was given if indicated based on TNM stage (T3-4 or N+). Chemotherapy consisted of cisplatin (100 mg/m² on days 1, 22, and 43 of treatment) or cetuximab (400 mg/m² initial dose, followed by a weekly dose of 250 mg/m²).

6.2.3 MR imaging and post-processing

All MR imaging was performed on a 1.5 T GE MR450w (GE, Waukesha, WI, USA) using the MR Radiation Oncology Suite coils (GE, Waukesha, WI, USA) with the patient immobilized in the radiotherapy treatment mask. The planning MRI protocol contained multi b-value DWI, a DWI scan with inverse phase encoding gradient polarity of only b=0 s/mm² for the purpose of distortion correction [47, 76], a T2-weighted (T2w) TSE and a T1-weighted (T1w) IDEAL [159]. Gadolinium-based contrast agent was administered before the start of the protocol. The multi b-value DWI scan (single-shot echo planar imaging, flip angle: 90 degrees TR: 6700 ms; TE: 81.8 ms; FOV: 26 x 26 cm; 4 mm slice thickness; 0.2 mm slice gap, 128 x 128 matrix, acceleration factor 2) consisted of 15 b-values (0, 10, 2x80, 130, 570, 2x770, 2x780, 790, and 4x1500 s/mm²) acquired in three orthogonal diffusion directions. These b-values are the result of a b-value optimization described in detail in previous work [121]. Distortion correction of the DWI was done with FSL topup [47, 76], based on the b=0 s/mm² images.

The full workflow of processing the DWI scans is depicted graphically in Figure 6.1. For conventional DWI, voxel-wise least square fitting was done with the mono-exponential model (Equation 6.1), and for NG-IVIM DWI according to the NG-IVIM model (Equation 6.2).

$$S_i = S_0 e^{-b_i ADC} \quad (6.1)$$

In Equation 6.1 S_i is the measured signal intensity at the corresponding b-value b_i , S_0 is the signal intensity at a b-value of 0 s/mm² and ADC is the apparent diffusion coefficient.

$$S_i = S_0 \left((1 - f) e^{-b_i D + \frac{1}{6} (b_i D)^2 K} + f e^{-b_i D^*} \right) \quad (6.2)$$

In Equation 6.2 S_i is the measured signal intensity at the corresponding b-value b_i , S_0 is the signal intensity at b-value of 0 s/mm², D is the diffusion coefficient, f is the perfusion fraction, D^* is the pseudo-diffusion coefficient, and K is the kurtosis.

All fitting was done within the tumor volume using all b-values except b=0 s/mm² with an in-house fitting algorithm employing a multiple starting point method written in MATLAB (MathWorks, Natick, MA, USA). For each voxel 1000 starting points were chosen between $0.25 \cdot 10^{-3}$ and $3.41 \cdot 10^{-3}$ mm²/s for ADC and D , 0.09 and 0.42 for f , $6.29 \cdot 10^{-3}$ and $23.39 \cdot 10^{-3}$ mm²/s for D^* and between 0.1 and 3 for K according to a Halton sequence [71]. These starting points form a set of low-discrepancy, pseudo-random vectors, and the ranges were chosen based on previous work [121] and visual inspection of the fit results. To avoid non-physiological results and extreme outliers, the following fitting constraints were used: 0 to $20 \cdot 10^{-3}$ mm²/s for ADC and D , 0 to 1 for f , 0 to $200 \cdot 10^{-3}$ mm²/s for D^* , and 0 to 5 for K .

Next, the average ADC , f , D^* , D , and K were calculated by averaging over all gross tumor volume (GTV) voxels.

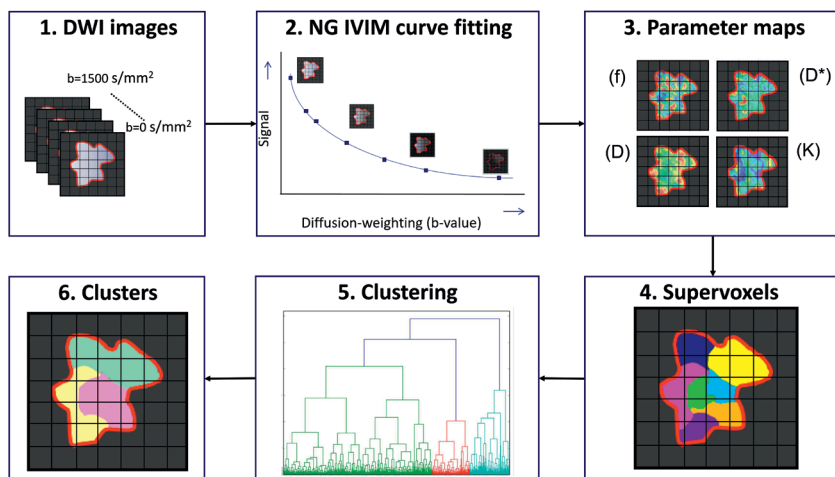


Figure 6.1. Graphical depiction of the workflow where the distortion corrected DWI images (1) are first fitted to the NG-IVIM model (2), which yields four parameter maps (3). Next, supervoxels are created based on the four parameter maps (4). Then all supervoxels from all patients are clustered (5) to gain insight into the prevalence of certain combinations of parameters (identifying certain phenotypes) (6) in different tumors.

6.2.4 Gross tumor volume delineation

The gross tumor volume (GTV) was delineated on the T2w images by an experienced radiation oncologist, with additional information from gadolinium-enhanced T1w images. Subsequently, the T2w image was rigidly (rotation and translation) registered to the distortion corrected $b=0 \text{ s/mm}^2$ image of the DWI scan for each patient, and the contours were propagated from the T2w image to the DWI. Next, the propagated GTV delineations on the DWI ($b=0 \text{ s/mm}^2$) were manually checked. If the shape of the pharynx deviated between the T2w and the $b=0 \text{ s/mm}^2$ scan, the voxels from the GTV located in air on the $b=0 \text{ s/mm}^2$ scan were excluded.

6.2.5 Assessing intra-tumor regions

In addition to the average DWI parameters value, intra-tumor parameter heterogeneity was investigated to identify tumor regions with similar phenotypes across different patients, using unsupervised clustering. Before applying the clustering methods over all the patients, noise was filtered out for each GTV by creating supervoxels using the Simple Linear Iterative Clustering (SLIC) algorithm [160]. A supervoxel can be seen as a union of adjacent voxels with similar normalized NG-IVIM parameters. The SLIC algorithm automatically

determines which voxels belong to which supervoxels, based on an average size set to 50 ± 10 voxels per supervoxel and a compactness set to 20.

Next, tumor regions with similar phenotypes across different patients were identified by clustering all supervoxels of all patients, using an agglomerative Ward clustering algorithm, based on the average normalized parameter values of the supervoxels [161]. This clustering approach was similar to Even et al. [162]. The optimal number of clusters (i.e., regions with similar phenotypes) was selected for the entire dataset (between two and ten) based on the Calinski-Harabasz index [163]. The number of clusters with the highest Calinski-Harabasz index was selected. Per tumor, the prevalence of each cluster (i.e., tumor region with distinct phenotype) was calculated by dividing the number of supervoxels belonging to that cluster by the total number of supervoxels of the tumor, which is referred to as the fractional contribution. Clustering was carried out with in-house software and the SciPy package (version 1.10.1) in python 3.8. For a more extensive explanation of the clustering procedure see Supporting Information section 6.6.1.

6.2.6 HPV typing

Immunohistochemical analysis was performed for p16^{INK4A}. Strong and diffuse nuclear and cytoplasmic immunostaining in more than 70% of the tumor cells was considered as p16 positive [164-166]. If HPV status was available, this was used instead of the p16 staining due to its lower false positive rate.

6.2.7 Response assessment

Patients were followed by the HN multidisciplinary team and response evaluation was performed by clinical examination and MR imaging, if indicated. Follow-up visits were bi-monthly for the first year following RT. Progressive disease within one year was defined as local disease, regional disease, distant metastasis, or any combination thereof present within one year after the end of radiotherapy. Complete response was defined as the absence of progressive disease.

6.2.8 Statistical analysis

The average *ADC*, *f*, *D**, *D*, *K*, and fractional cluster contribution were compared between HPV-positive and HPV-negative tumors using Wilcoxon rank sum tests. The parameter values (*f*, *D**, *D*, *K*) per cluster were compared using Kruskal-Wallis and Dunn's post hoc tests. The average *ADC*, *f*, *D**, *D*, *K*, fractional cluster contribution, and response within one year were compared with Wilcoxon rank sum tests in subgroup analyses of HPV-negative and HPV-positive patients separately. To investigate a possible confounding effect of T and N stage on response, the correlation between tumor volume, T and N stage, and *ADC*,

f , D^* , D , and K was investigated using Spearman correlation (r_s) in subgroup analysis of HPV-negative and HPV-positive patients separately. A p-value of $p < 0.05$ was considered statistically significant. No correction for multiple testing was used. All statistical analyses, including clustering, were carried out with the SciPy package (version 1.10.1) in python 3.8.

6.3 Results

6.3.1 Patient characteristics

Two patients had to be excluded due to a poor fit of the NG-IVIM model caused by the signal being insufficient at $b=1500$ s/mm². In total 36 patients remained, of which 18 were HPV-positive and 18 HPV-negative. The average GTV volume was 16.3 cc (range 0.9 – 106.7 cc). Table 6.1 shows the patient characteristics and response one year post (chemo)radiotherapy per HPV status. Table 6.S1 shows the patient characteristics in the HPV-negative group per treatment outcome (complete response or progressive disease).

Table 6.1. Patient characteristics. Tumor staging was done according to TNM classification, edition 8.

	Total	HPV-positive	HPV-negative
N	36	18	18
Age [years] (mean ± SD)	62±8	61±7	62±9
Sex			
Male	26	13	13
Female	10	5	5
T Stage			
T1-2	22	13	9
T3-4	14	5	9
N stage			
N0	13	4	9
N+	23	14	9
M stage			
M0	36	18	18
M+	0	0	0
Tumor volume [cc] (mean ± SD)	16±22	13±12	20±28
Smoking at start RT			
Yes	20	7	13
No	16	11	5
Never smokers	5	5	0

Table 6.1. *Continued.*

	Total	HPV-positive	HPV-negative
Former smokers	11	6	5
Radiotherapy			
Photons	26	14	12
5 fr/week	14	9	5
6 fr/week	12	5	7
Protons	10	4	6
5 fr/week	6	4	2
6 fr/week	4	0	4
Chemotherapy			
Yes	26	15	11
Cisplatin	17	12	5
Cetuximab	9	3	6
No	10	3	7
Response one-year post-RT			
Complete response	28	16	12
Progressive disease	8	2	6
Local failure	5	2	3
Regional failure	3	1	2
Distant metastasis	3	0	3

6.3.2 Differences between HPV-positive and HPV-negative patients

Figure 6.2 shows boxplots of the distributions of the *ADC* and *NG-IVIM* parameters over the patients stratified by HPV status. The mean *ADC* (\pm SD) of the HPV-negative group was $1.4 \pm 0.4 \cdot 10^{-3}$ mm²/s, which was significantly higher ($p=0.018$) than that of the HPV-positive group at $1.1 \pm 0.2 \cdot 10^{-3}$ mm²/s. In the HPV-negative group, the *NG-IVIM* parameters *D* and *f* were significantly higher than in the HPV-positive group: for *D* $1.4 \pm 0.2 \cdot 10^{-3}$ mm²/s versus $1.2 \pm 0.2 \cdot 10^{-3}$ mm²/s ($p=0.031$), for *f* 0.24 ± 0.08 versus 0.19 ± 0.06 ($p=0.037$). The *D** and *K* were significantly lower in the HPV-negative group compared to the HPV-positive group: for *D** $2.3 \pm 0.5 \cdot 10^{-2}$ mm²/s versus $2.5 \pm 0.3 \cdot 10^{-2}$ mm²/s ($p=0.016$), for *K* 0.8 ± 0.2 versus 1.0 ± 0.3 ($p=0.034$).

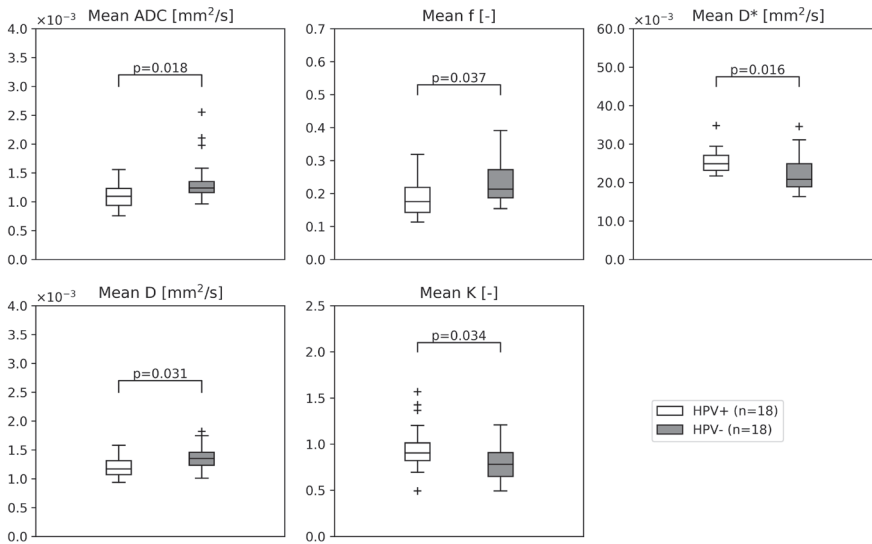


Figure 6.2. Boxplots of the ADC and the NG-IVIM parameters in which the white boxplot depicts parameter values for the HPV-positive patients and the gray boxplot for the HPV-negative patients. The horizontal line represents the median, and the box represents the 25th to 75th percentile.

6.3.3 Intra-tumor regions

In order to assess differences in intra-tumor regions between HPV-positive and HPV-negative patients, the tumors were divided into 4 to 483 supervoxels per tumor depending on tumor size, with a mean of 75 supervoxels per tumor. The optimal number of clusters according to the Calinski-Harabasz index was three. The average parameter values of these three clusters are schematically depicted in Figure 6.3a. For all parameters (f , D^* , D , and K), the parameter values were significantly different between all clusters ($p < 0.001$).

Cluster one showed a significantly higher contribution in HPV-negative tumors compared to HPV-positive tumors ($p = 0.033$), while cluster three showed a trend towards lower contribution in HPV-negative tumors compared to HPV-positive tumors ($p = 0.054$) (Figure 6.3c).

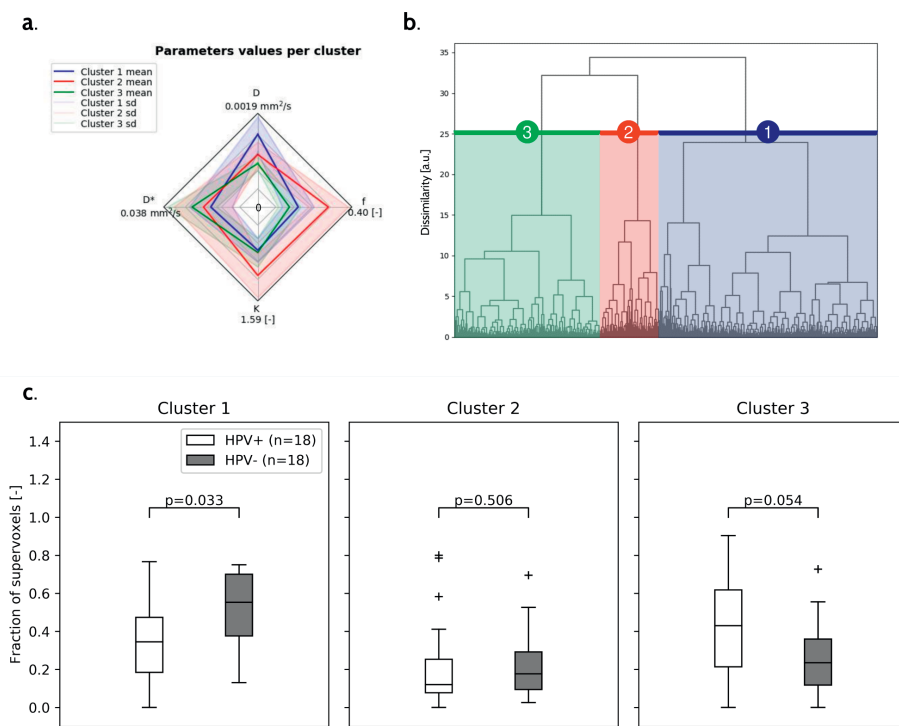


Figure 6.3. (a) A spider web plot is shown with the average value of each parameter for the three clusters. The center of the spider web plot represents the value 0. (b) The clustering tree is shown with the corresponding cluster numbers referred to in (c). (c) Boxplots of the fractional contribution of each of the clusters for HPV-positive and HPV-negative patients. The p-values are not corrected for multiple testing. The horizontal line represents the median, and the box represents the 25th to 75th percentile.

6.3.4 Tumor response, HPV status, and conventional DWI and NG-IVIM DWI parameters

Due to the rare occurrence of progressive disease in HPV-positive patients within one year (2 out of 18 patients), subgroup analysis was only done for the HPV-negative patient group, in which 6 out of 18 had progressive disease within one year after treatment. The mean D was significantly lower in HPV-negative patients with progressive disease at $1.2 \pm 0.1 \cdot 10^{-3} \text{ mm}^2/\text{s}$ compared to HPV-negative patients with a complete response at $1.4 \pm 0.2 \cdot 10^{-3} \text{ mm}^2/\text{s}$ ($p=0.015$). ADC , f , D^* , and K did not show a significant difference between complete response and progressive disease (Figure 6.4). No significant correlation between T stage or N stage and the DWI parameters ADC , D , f , D^* , and K was found in the HPV-negative subgroup ($p\text{-value} > 0.05$ and r_s ranged from -0.45 to 0.30). This suggests that T stage and N stage are not confounding factors

for the relation between DWI parameters and response. For tumor volume, a significant correlation was found with f ($r_s = -0.61$, $p = 0.007$), but not for the other parameters. Therefore, this suggests that tumor volume is not a confounding factor for the relation between ADC , D^* , D , and K and response.

Cluster one showed a significantly higher contribution in HPV-negative patients with a complete response compared to progressive disease ($p = 0.015$), while cluster three showed a significantly lower contribution in HPV-negative patients with a complete response compared to progressive disease ($p = 0.009$) (Figure 6.5).

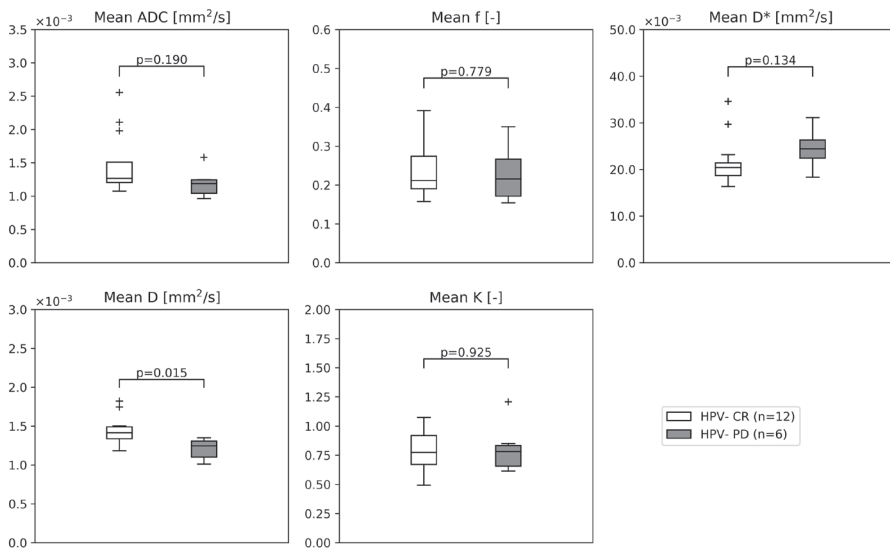


Figure 6.4. Boxplots of the ADC and the NG-IVIM parameters; the white boxplots depict parameter values for the HPV-negative patients with a complete response (CR) and the gray boxplots for the HPV-negative patients with progressive disease (PD). The horizontal line represents the median, and the box represents the 25th to 75th percentile.

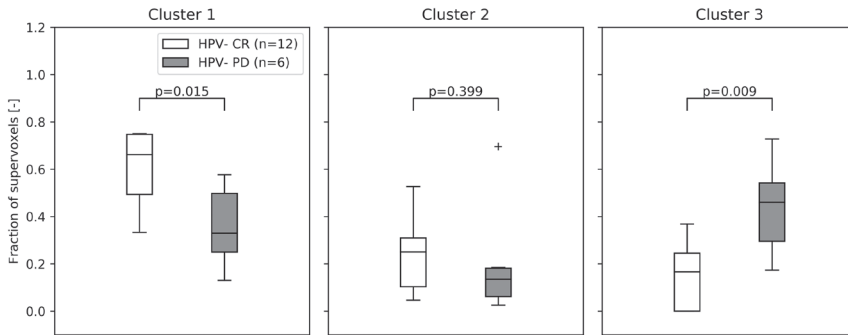


Figure 6.5. Boxplots of the fractional contribution of each of the clusters for HPV-negative complete responders (CR) and HPV-negative patients with progressive disease (PD) within one year. The p-values are not corrected for multiple testing. The horizontal line represents the median, and the box represents the 25th to 75th percentile.

6.4 Discussion

In this study, we describe the first clinical application of our recently optimized NG-IVIM acquisition for a group of 36 oropharyngeal tumor patients. NG-IVIM enables simultaneous assessment of *inter*-cellular diffusion (similar to conventional DWI), microvascular perfusion, and *intra*-cellular diffusion.

We found that the *ADC* and NG-IVIM parameters f , D^* , D , and K were related to HPV status. HPV-negative patients had a higher *ADC* and NG-IVIM parameter D than HPV-positive patients, which is in line with existing literature [11, 156, 158, 167-177]. Since HPV-negative patients generally have a worse response to treatment than HPV-positive patients, that finding by itself would suggest that a high *ADC* is related to a poorer response, which also has been reported before in studies about pretreatment DWI that did not correct for HPV status [37, 178, 179]. However, it turned out that in our cohort of HPV-negative patients, the opposite correlation was found, namely that a lower NG-IVIM parameter D was related to poorer response. No relation between *ADC* and response was observed. In other words, our results could suggest that *ADC* is a surrogate for HPV status and is not related to response, while NG-IVIM parameter D was related to response in the HPV-negative group. Also, it implies that NG-IVIM response analyses should be performed separately for HPV-positive and HPV-negative patients. The finding that NG-IVIM parameter D was related to response, while *ADC* was not, could be a concrete indication of the added value of NG-IVIM compared to conventional DWI.

Martens et al. [155] did correct for HPV status using a multimodality CoxBoost regression analysis, but none of the pretreatment IVIM parameters nor *ADC* were significant predictors for locoregional failure. Yet, when assessing overall survival, Martens et al. [155] did find both pretreatment high mean *ADC* and low *D** as predictive factors for poor overall survival, while Ravanelli et al. [154] did not find any correlation between pretreatment *ADC* and overall survival when analyzing the HPV-positive and HPV-negative group separately. Similarly, Connor et al. [176] did not find a correlation between pretreatment *ADC* and disease-free survival after 2 years in the subgroup containing HPV-positive OPSCC only, nor in the subgroup of the other head and neck carcinomas (including HPV-negative OPSCC). Therefore, further research is needed to determine the prognostic value of DWI parameters within HPV subgroups.

Note that the b-value set used in this study resulted from a recent study [121] where we optimized the b-values for NG-IVIM to yield maximum parameter estimation precision given the number of b-values. That study was performed specifically as a step towards standardization, because no consensus on the exact b-value set exists in literature, and the b-values impact the parameter values that are estimated. Since the optimized b-value set differs from the (variety of) b-value sets used in the mentioned studies, and also in literature either the *ADC* model or IVIM model are used, differences between literature and the current study could partly be explained by differences in the combination of the used model and the b-value sets. For example, the *ADC* may be biased substantially depending on whether b-values in the perfusion range (0–200 s/mm²) and restricted diffusion range (>800 s/mm²) are used or not. Since *f*, *D**, and *K* did not differ between progressive disease and complete responders, and the *ADC* we used also includes perfusion and restricted diffusion effects, it may have outweighed the fact that *D* did differ between progressive disease and complete response. In turn leading to the result that *ADC* did not significantly differ between progressive disease and complete response.

ADC showed a stronger difference between HPV-positive and HPV-negative tumors than *D*. Since all b-values were used to fit the mono-exponential model for the *ADC*, the difference between *ADC* and *D* most likely reflects the inclusion of the perfusion and/or kurtosis effects in the *ADC*. This effect might also be reflected in the larger ranges of *ADC* values compared to *D* values. The higher *ADC* and *D* in HPV-negative tumors compared to HPV-positive tumors could be due to the fact that HPV-negative tumors tend to have variable cellularity and high stromal content, whereas HPV-positive tumors tend to have back-to-back densely packed cells and less tumor stromal component [11].

NG-IVIM parameter K quantifies the non-Gaussian diffusion behavior of water molecules when diffusion is restricted by cell membranes or other microstructural components [42]. This means that tissue with smaller cells has a higher K . The lower average K found in HPV-negative tumors compared to HPV-positive tumors could be caused by the fact that HPV-positive tumors have more cells with basaloid appearance (which are generally smaller than cells without basaloid appearance) and more infiltration of lymphocytes [180]. This highlights the benefit of extending the conventional DWI model not only to the IVIM DWI model, but to the NG-IVIM DWI model, as both the conventional and IVIM DWI models do not incorporate the non-Gaussian diffusion behavior of water molecules.

We also found differences in f and D^* between HPV-positive and HPV-negative OPSCC. The higher average f and lower average D^* in HPV-negative tumors compared to HPV-positive tumors suggest that even though there is more blood volume, the blood velocity is lower in HPV-negative tumors. These trends in perfusion parameters might suggest that HPV-negative OPSCC has less functional vasculature. This hypothesis is supported by the study of Hanns et al. [181] that showed a lower density of neo-blood vessels, more hypoxic tumor areas, and higher mRNA expression of hypoxia-responsive genes in HPV-negative tumors compared to HPV-positive HN tumors. However, Vidiri et al. [173] did not find any significant differences in f and D^* for HPV-negative versus HPV-positive OPSCC, and contradicting literature can be found about the vascularization of HPV-negative and HPV-positive tumors [181, 182].

In addition to the average parameter value, intra-tumor heterogeneity was investigated by analyzing regions with similar NG-IVIM parameter values, using unsupervised clustering. One cluster had a significantly higher presence in HPV-negative tumors and one cluster had a higher, albeit not significantly, presence in HPV-positive tumors. This suggests that HPV-negative tumors often have regions with a relatively high D (related to a high amount of stroma) and HPV-positive tumors often have regions with a relatively high D^* (related to high blood velocity).

DWI in the HN is prone to motion. While most head motion is mitigated as patients are scanned in the immobilization mask, misalignment between the $b=0$ s/mm² and higher b -values due to swallowing and/or coughing may occur. In general, swallowing/coughing artifacts affect only one b -value, so unless a patient was coughing or swallowing excessively, the effect on the parameter values would be minimal.

In this study, we focused on the possible value of pretreatment DWI for response assessment of the primary tumor. Several prior studies indicated that obtaining DWI during treatment [36, 73, 151, 183] could be interesting for response assessment as well. Additionally, assessing lymph nodes could yield information on response as well and could be of interest for future studies [184].

This study has some limitations. First, we used p16 status as a proxy for HPV status. However, p16 is known to be false positive in around 5–20% of the cases [185]. Therefore, it is likely that some patients defined as HPV-positive in this study were false positives. Second, due to the small sample size, relatively short follow-up, and the single-center nature of the study, more research should be performed to ensure the results are generalizable. Moreover, due to the relatively small sample size, we did not correct for multiple testing. Finally, a limitation of unsupervised clustering is that adding or removing patients could result in slightly different clusters than presented here.

6.5 Conclusion

In this study, our recently optimized NG-IVIM was applied for the first time to a group of 36 OPSCC patients. Differences in ADC as well as all NG-IVIM parameters and in cluster analyses were found between HPV-positive and HPV-negative tumors. In a subgroup analysis of only HPV-negative patients, we found that D negatively correlated with progressive disease, which contradicts current literature relating ADC and D to progressive disease without correcting for HPV status. This suggests that ADC and D estimated in those studies could potentially be a surrogate for HPV status instead of a response predictor. Therefore, HPV status should be corrected for when assessing the predictive value of DWI. We found no correlation between response and ADC , indicating the potential added value of the more elaborate NG-IVIM model compared to conventional DWI.

6.6 Supporting information

6.6.1 Assessing intra-tumor regions

Filtering out noise

First, noise was filtered out by creating supervoxels using the Simple Linear Iterative Clustering (SLIC) algorithm [160] for each tumor separately. Supervoxels are combinations of several voxels that are located close together and have similar image intensity across multiple parameter maps. To ensure equal weighting of the parameters (f , D^* , D , and K), each parameter was normalized separately before running the supervoxel algorithm. The normalization was

based on the average of the mean parameter value per GTV (\bar{p}), and the average of the standard deviation per GTV ($\overline{p_{std}}$) over all patients according to Equation 6.S1 in which $p_{normalized}$ and p are the normalized and not normalized parameter values of a voxel respectively. Within the SLIC supervoxel algorithm, the average size was set to 50 ± 10 voxels and the compactness to 20.

$$p_{normalized} = \frac{p - \bar{p}}{\overline{p_{std}}} \quad (6.S1)$$

Clustering algorithm

After creating supervoxels, the average normalized parameter values were calculated per supervoxel as input for the clustering algorithm. The resulting supervoxels of all patients were combined into one dataset. The dataset was clustered using an agglomerative Ward clustering algorithm based on the average normalized parameter values [161]. This algorithm starts by assigning a cluster to each individual supervoxel. Subsequently, an iterative process started in which two clusters were combined into new clusters until only one cluster was left. The Ward method was used to decide which clusters were combined per iteration step. This entailed that the variance per cluster was calculated as well as the variance of each possible newly combined cluster. The newly combined cluster causing the smallest rise in total variance was selected.

Selecting the optimal number of clusters

After creating this cluster hierarchy, we had to choose which number of clusters would be optimal for our analysis. When using too many clusters, clusters can be very similar and might not relevantly differ from each other. On the other hand, using too few clusters can cause high intra-cluster variances and thereby erase relevant differences. Therefore, it was important to have a trade-off between low intra-cluster and high inter-cluster variation. The Calinski-Harabasz index is a measure that penalizes intra-cluster variability and rewards inter-cluster variability (Equation 6.S2) [163].

$$CH = \left[\frac{\sum_{k=1}^K n_k \|c_k - c\|^2}{K - 1} \right] / \left[\frac{\sum_{k=1}^K \sum_{i=1}^{n_k} \|d_i - c_k\|^2}{N - K} \right] \quad (6.S2)$$

In this formula, K is the number of clusters on the dataset $D = [d_1, d_2, \dots, d_N]$. n_k and c_k are the number of points and centroid of the k^{th} cluster respectively. c is the global centroid and N is the total number of data points. The cluster number with the highest Calinski-Harabasz index was selected. In our case, this was three (Figure 6.S1).

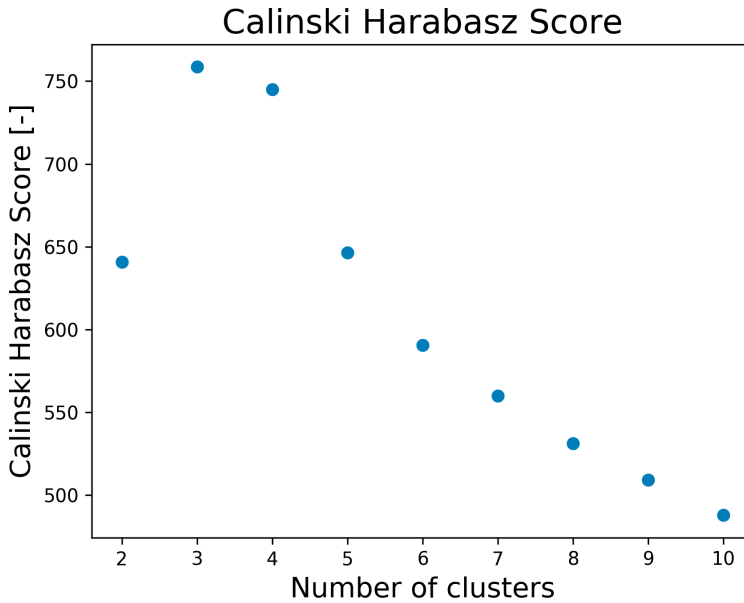


Figure 6.S1. An overview of the Calinski–Harabasz index for the different number of clusters. The higher the Calinski–Harabasz the better the clusters describe the variation in the data in terms of minimal intra-cluster variation and maximum inter-cluster variation.

Cluster prevalence

Per tumor, the prevalence of each cluster was calculated by dividing the number of supervoxels belonging to a cluster by the total number of supervoxels of the tumor (the fractional contribution).

6.6.2 Patient characteristics per response group

Table 6.S1. Patient characteristics of HPV-negative patients per response group. Tumor staging was done according to TNM classification, edition 8.

	HPV-negative	Complete response	Progressive diseases
N	18	12	6
Age [years] (mean ± SD)	62±9	60±8	67±8
Sex			
Male	13	8	5
Female	5	4	1

Table 6.S1. *Continued.*

	HPV-negative	Complete response	Progressive diseases
T Stage			
T1-2	9	7	2
T3-4	9	5	4
N stage			
N0	9	9	0
N+	9	3	6
M stage			
M0	18	12	6
M+	0	0	0
Tumor volume [cc] (mean ± SD)	20±28	18±22	25±36
Smoking at start RT			
Yes	13	9	4
No	5	3	2
Never smokers	0	0	0
Former smokers	5	3	2
Radiotherapy			
Photons	12	11	1
5 fr/week	5	4	1
6 fr/week	7	7	0
Protons	6	1	5
5 fr/week	2	0	2
6 fr/week	4	1	3
Chemotherapy			
Yes	11	7	4
Cisplatin	5	4	1
Cetuximab	6	3	3
No	7	5	2

Part Two

7

Accounting for fractionation and heterogeneous dose distributions in the modeling of osteoradionecrosis in oropharyngeal carcinoma treatment

Gerda M. Verduijn, **Nienke D. Sijtsema**,
Yvette van Norden, Wilma D. Heemsbergen,
Hetty Mast, Aniel Sewnaik, Denzel Chin,
Sarah Baker, Marta E. Capala,
Aad van der Lugt, Esther van Meerten,
Mischa S. Hoogeman, Steven F. Petit

*This chapter is based on Verduijn GM,
Sijtsema ND, van Norden Y,
Heemsbergen WD, Mast H, Sewnaik A, et al.
Accounting for fractionation and heterogeneous
dose distributions in the modelling of
osteoradionecrosis in oropharyngeal carcinoma
treatment. Radiother Oncol. 2023;188:109889.*

Abstract

Background and purpose

Osteoradionecrosis (ORN) of the mandible is a severe complication following radiotherapy (RT). With a renewed interest in hypofractionation for head and neck radiotherapy, more information concerning ORN development after high fraction doses is important. The aim of this explorative study was to develop a model for ORN risk prediction applicable across different fractionation schemes using Equivalent Uniform Doses (EUD).

Materials and methods

We performed a retrospective cohort study in 334 oropharyngeal squamous cell carcinoma (OPSCC) patients treated with either a hypofractionated Stereotactic Body Radiation Therapy (HF-SBRT) boost or conventional Intensity Modulated Radiation Therapy (IMRT). ORN was scored with the CTCAE v5.0. HF-SBRT and IMRT dose distributions were converted into equivalent dose in 2 Gy fractions ($\alpha/\beta = 0.85$ Gy) and analyzed using EUD. The parameter a that led to an EUD that best discriminated patients with and without grade ≥ 2 ORN was selected. Patient and treatment-related risk factors of ORN were analyzed with uni- and multivariable regression analysis.

Results

A total of 32 patients (9.6%) developed ORN grade ≥ 2 . An EUD ($a = 8$) best discriminated between ORN and non-ORN (AUC = 0.71). In multivariable regression, pre-RT extractions (SHR = 2.34; $p = 0.012$), mandibular volume (SHR = 1.04; $p = 0.003$), and the EUD ($a = 8$) (SHR = 1.14; $p < 0.001$) were significantly associated with ORN.

Conclusion

Risk models for ORN based on conventional DVH parameters cannot be directly applied to HF-SBRT fractionation schemes and dose distributions. However, after correcting for fractionation and nonuniform dose distributions using EUD, a single model can distinguish between ORN and non-ORN after conventionally fractionated radiotherapy and hypofractionated boost treatments.

7.1 Introduction

The majority of patients with oropharyngeal squamous cell carcinoma (OPSCC) are treated with radiotherapy (RT) with or without systemic therapy. A serious late complication is osteoradionecrosis (ORN) of the jaws, which develops most often in the mandible. ORN may cause serious morbidity such as pain, deformity, and even pathologic fractures that need extensive reconstruction [186–188]. The incidence of ORN in the Intensity Modulated Radiation Therapy (IMRT) era varies considerably between 2% and 12% [54, 189–192].

Different risk factors for ORN have been suggested, such as smoking, dental extractions, and the radiation dose to the mandible [54, 55, 189, 191–195]. However, although radiation dose is an important risk factor, specific dosimetric parameters to guide radiotherapy planning are limited [54, 55, 189, 191–198] and are all commonly based on studies with limited variation in the fraction dose and dose distribution due to stringent constraints on dose uniformity in the tumor. Therefore, existing ORN risk models cannot simply be generalized to other types of dose distributions, for example with a more heterogeneous dose distribution in the tumor, or to higher fraction doses.

With the upcoming interest in hypofractionated head and neck radiotherapy and more heterogeneous dose distributions [16, 199–201], there is an urgent need for a better understanding of the influence of alternative fractionation schemes on the risk of ORN. For ORN in particular, this is important as a low α/β for the mandible has been suggested, indicating an increased sensitivity to a high fraction dose [202].

Therefore, the aim of this study was to explore whether a single model that uses EUD after conversion of the dose distributions to an equivalent dose in 2 Gy fractions, can describe ORN for different fractionation schemes and dose heterogeneities. To this end, we included in the modeling OPSCC patients that were treated with IMRT (with 23 fractions of 2 Gy) followed by either a hypofractionated Stereotactic Body Radiation Therapy (HF-SBRT) boost (with 3 fractions of 5.5 Gy), or a conventional IMRT boost (with 12 fractions of 2 Gy).

7.2 Materials and methods

7.2.1 Patients

The study was approved by the Medical Ethical Committee of Erasmus Medical Center (EMC17404). The records of all patients with OPSCC treated at the Department of Radiotherapy at the Erasmus Medical Center between January 2009 and May 2016 who survived at least one year following radiotherapy

were reviewed. Eligibility criteria were OPSCC patients treated with curative (chemo)radiation. Exclusion criteria were diagnosis with another primary malignancy within six months, previous oropharyngeal cancer, or previous head and neck radiotherapy. Patients lost to follow-up, and those with tumor progression within six months were also excluded.

7.2.2 Treatment and follow-up

All patients were discussed during a weekly multidisciplinary tumor board and treated with (chemo)radiation according to standard clinical protocols. cT1-smallT3N0-2cM0 tumors were treated with 46 Gy IMRT to the macroscopic tumor and elective lymph node regions with five to six fractions of 2 Gy per week. This was followed by a HF-SBRT boost to the primary tumor of 16.5 Gy in three daily fractions of 5.5 Gy with the Cyberknife system (Accuray Inc., Sunnyvale, CA, USA). After the HF-SBRT boost, neck dissection was performed in case of initial N+ disease (HF-SBRT boost group). Large cT3-T4N0-2M0 and any cT1-4N3M0 tumors, on the other hand, were treated with 70 Gy IMRT to the macroscopic tumor and pathologic lymph nodes with five to six fractions of 2 Gy per week and 46 Gy of IMRT to the elective lymph node regions (IMRT boost group). If indicated (cT3-T4 or N+ and age \leq 70), cisplatin (100 mg/m² on days 1, 22, and 43) or cetuximab (400mg/m² initial dose, followed by a weekly dose of 250 mg/m²) was added to the treatment. All patients were immobilized with a thermoplastic mask. For the IMRT treatment, the target coverage objective was that at least 98% of the PTV must receive at least 95% of the prescription dose (V95 > 98%). For the HF-SBRT boost, 16.5 Gy was prescribed to the 80% isodose line, where the maximum dose was normalized at 100%. For both fractionation regimens, the dose constraints for the total plan (EQD2 with $\alpha/\beta = 3$ Gy) were as follows: spinal cord Dmax < 50 Gy, brainstem Dmax < 60 Gy (both hard planning constraints), parotid glands Dmean < 26 Gy, submandibular glands Dmean < 39 Gy, oral cavity Dmean < 50 Gy, and constrictor muscles Dmean < 55 Gy (when achievable). At the time of treatment of this patient group, no constraints were used for the mandible.

Patients were routinely seen at the Department of Oral and Maxillofacial Surgery before start of radiotherapy, and dental extractions were performed following national guidelines [203]. During radiotherapy treatment, dentulous patients were seen weekly by a dental hygienist and received fluoride treatment. Follow-up visits were planned every two months for the first year following RT by the head and neck multidisciplinary team. The frequency gradually decreased to every six months for a minimum of five years. In case of suspicion or diagnosis of ORN, patients were referred to the ORN outpatient clinic at the Department of Oral and Maxillofacial Surgery.

7.2.3 Toxicity scoring

ORN was defined as clinically exposed bone following radiotherapy, without evidence of recurrent or residual tumor existing for at least three months. This was determined by physical examination by an experienced head and neck radiation oncologist and oral and maxillofacial surgeon, and by radiological examination in all cases. In case of doubt of tumor recurrence, histological examination was performed. ORN as late toxicity (> 90 days after completion of RT) was retrospectively scored for all patients with the CTCAE v5.0 grading system (Supporting information Table 7.S1).

7.2.4 Dosimetric assessment

All dose distributions for both treatment groups were converted into the equivalent dose in 2 Gy fractions (EQD2). For the mandible and parotid glands, we used an $\alpha/\beta = 0.85$ Gy [202] and an $\alpha/\beta = 3$ Gy, respectively. For patients receiving a HF-SBRT boost, the 46 Gy IMRT plan and the 16.5 Gy HF-SBRT boost were based on different planning CT scans. Before dose summation of the two plans, the HF-SBRT CT scan was registered to the IMRT CT scan, using rigid registration focused on the mandible. Note that due to the rigidity of the mandible, rigid image registration was appropriate to derive the sum DVH of the mandible.

The following dose metrics were determined for the mandible for the summed plan: maximum dose (Dmax), mean dose (Dmean), and the volume of the mandible in cc receiving 10 Gy (V10Gy) up to 105 Gy (V105Gy) in 5 Gy intervals. The parotids Dmean (both glands combined) was also calculated, because hyposalivation places patients at risk for dental toxicity and subsequent ORN [204]. For calculation of the summed Dmean of the parotid glands no registration was required.

Due to the accepted target dose inhomogeneity of SBRT treatments and the difference in fraction dose, the different treatments (HF-SBRT vs. IMRT) led to very different dose volume histograms (DVH) for the mandible. To account for this difference, the DVHs of the mandible were converted to EUD values with our in-house developed Matterhorn framework, with different values for α ranging between 1 and 20, according to

$$EUD(\alpha) = \left(\sum_i v_i D_i^\alpha \right)^{\frac{1}{\alpha}} \quad (7.1)$$

with v_i the volume of a voxel, and D_i the dose to the voxel.

The EUD provides a single metric for reporting non-uniform dose distributions [205] and therefore it allows inclusion of different types of dose distributions in the modeling of ORN. It incorporates a specific volume effect with the EUD parameter α . When $\alpha = 1$, the EUD equals the mean dose. Increasing EUD parameter α , increases the importance of volumes receiving high doses compared to the volumes receiving low doses.

Since the different Dmax, VGy, and EUD values are expected to be strongly correlated, the following procedure was used to detect the most suited dose metric for the statistical analysis. For each dose metric, the receiver operator characteristics (ROC) curves were calculated to determine how well the different dose parameters could discriminate patients with and without ORN. The dose metric that led to the highest area under the curve (AUC) was used for the subsequent statistical analysis. These AUC values are purely descriptive; no statistical tests were applied in this preselection step.

7.2.5 Statistical analyses

Statistical analyses were performed using STATA software (Release 15, College Station, TX: StataCorp LLC). p-values $< .05$ were considered statistically significant. Follow-up time for toxicity was calculated from the last day of radiotherapy. Only ORN toxicity before the time of disease recurrence was taken into account. As an alternative to the Cox models, Fine & Gray models were used to account for the competing event of death. For the event of interest, occurrence of ORN grade ≥ 2 , subhazard ratios (SHR), as well as p-values are presented. Risk factors for ORN grade ≥ 2 evaluated in univariable regression models with a p-value < 0.2 were tested in multivariable models using the backward selection method. For the dose parameters, only the EUD which best discriminated patients with and without ORN was incorporated in the backward selection to avoid strongly correlated covariates in the multivariable model. Risk factors assessed included: sex, WHO performance status (0 vs. 1-2), alcohol consumption (no or previous vs. current), smoking (no or previous vs. current), pre-RT extractions (no or edentulous vs. partly or completely), time between extractions and RT (days), T stage (T3-4 vs. T1-2), N stage (N2-3 vs. N0-1), tumor subsite (tonsil vs. other), systemic therapy (yes vs. no), RT treatment scheme (IMRT boost vs. HF-SBRT boost), accelerated RT (yes vs. no), neck dissection (yes vs. no), mandibular volume in cc, and the dosimetric parameters with the highest AUC (as described in the previous section).

7.3 Results

The number of patients that met the inclusion criteria was 334 of which 333 patients completed radiation therapy. In one patient radiation therapy was

canceled after 60 Gy due to age and comorbidity. This was taken into account in the dosimetric analyses as well. Patient, tumor, and treatment characteristics are shown in Table 7.1. Median follow-up for patients without ORN and still alive at the date of last contact was 55 months (range 8–119).

A total of 32 patients (9.6%) developed ORN grade ≥ 2 (5.4% grade 2, 3.9% grade 3, 0% grade 4, 0.3% grade 5). Median time to onset of ORN was 19.5 months with a prevalence of ORN grade ≥ 2 of 7% at 3 years and 10% at 5 years (Figure 7.1). In the HF-SBRT boost group, the crude incidence of ORN grade ≥ 2 was 9.3% (4.9% grade 2, 3.8% grade 3, 0% grade 4, 0.5% grade 5). In the IMRT boost group, the crude incidence of ORN was 9.9% (5.9% grade 2, 5.9% grade 3).

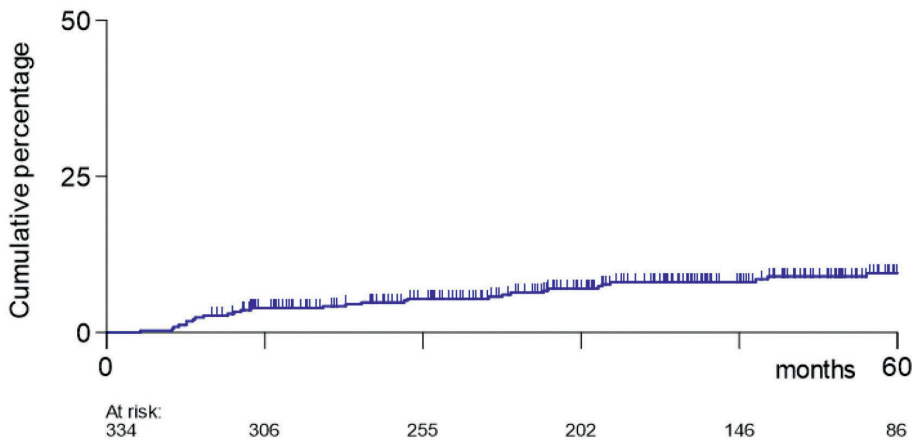


Figure 7.1. Kaplan-Meier plot showing the cumulative incidence of ORN \geq grade 2 (CTCAE v5.0). Abbreviations: ORN = Osteoradionecrosis; CTCAE = Common Toxicity Criteria of Adverse Events.

For 325 patients (94%) dosimetric data could be retrieved. The DVHs of the mandible of all these patients are shown in Figure 7.2. DVHs of the IMRT-only treatment are characterized by relatively large volumes receiving intermediate dose levels of 30–40 Gy (Figure 7.2a and 7.2c), while the HF-SBRT treatments have DVHs with long tails towards higher doses of more than 80 Gy (Figure 7.2b and 7.2d). For an overview of dose statistics between the IMRT and HF-SBRT boost group, see Supporting Information Table 7.S2.

Table 7.1. Patients and treatment characteristics of the total group and per boost type. Percentages are shown in brackets. Abbreviations: HF-SBRT = Hypofractionated Stereotactic Body Radiation Therapy; IMRT = Intensity Modulated Radiation Therapy; ORN = osteoradionecrosis; SD = Standard Deviation; WHO = World Health Organization; HPV = human papillomavirus.

Characteristics	All patients N = 334 (%)			HF-SBRT N = 182 (%)			IMRT N = 152 (%)		
	Non ORN	ORN	Total	Non ORN	ORN	Total	Non ORN	ORN	Total
ORN grade ≥ 2	302 (90)	32 (10)		165 (91)	17 (9)		137 (90)	15 (10)	
Age (mean; SD)	61; 8.7	62; 7.2	61; 8.5	61; 8.5	59; 7.2	61; 8.4	61; 8.9	65; 6.1	62; 8.7
Sex									
Male	199 (66)	21 (66)	220 (66)	101 (61)	11 (65)	112 (62)	98 (72)	10 (67)	108 (71)
Female	103 (34)	11 (34)	114 (34)	64 (39)	6 (35)	70 (38)	39 (28)	5 (33)	44 (29)
WHO performance									
0	175 (58)	19 (59)	194 (58)	130 (79)	15 (88)	145 (80)	45 (33)	4 (27)	49 (32)
1-2	127 (42)	13 (41)	140 (42)	35 (21)	2 (12)	37 (20)	92 (67)	11 (73)	103 (68)
Alcohol									
No/ previous	85 (28)	7 (22)	92 (28)	40 (24)	3 (18)	43 (24)	45 (33)	4 (27)	49 (32)
Current	217 (72)	25 (78)	242 (72)	125 (76)	14 (82)	139 (76)	92 (67)	11 (73)	103 (68)
Smoking									
No/ previous	146 (48)	11 (34)	157 (47)	78 (47)	7 (41)	85 (47)	68 (50)	4 (27)	72 (47)
Current	156 (52)	21 (66)	177 (53)	87 (53)	10 (59)	97 (53)	69 (50)	11 (73)	80 (53)
Pre-RT extractions									
No/ edentulous	195 (65)	13 (41)	208 (62)	113 (68)	6 (35)	119 (65)	82 (60)	7 (47)	89 (59)
Partly/ complete	107 (35)	19 (59)	126 (38)	52 (32)	11 (65)	63 (35)	55 (40)	8 (53)	63 (41)
Time (days) extractions – RT (median)	23	18	21	19	17	19	24	22	24

Table 7.1. Continued.

Characteristics	All patients N = 334 (%)			HF-SBRT N = 182 (%)			IMRT N = 152 (%)		
	Non ORN	ORN	Total	Non ORN	ORN	Total	Non ORN	ORN	Total
T stage									
T1-T2	183 (61)	18 (56)	201 (60)	142 (86)	14 (82)	156 (86)	41 (30)	4 (25)	45 (30)
T3-T4	119 (39)	14 (44)	133 (40)	23 (14)	3 (18)	26 (14)	96 (70)	11 (75)	107 (70)
N stage									
N0-N1	165 (55)	18 (56)	183 (55)	104 (63)	12 (71)	116 (64)	61(45)	6 (40)	67 (44)
N2-N3	137 (45)	14 (44)	151 (45)	61 (37)	5 (29)	66 (36)	76 (55)	9 (60)	85 (56)
Tumor subsite									
Tonsil	147 (49)	16 (50)	163 (49)	94 (57)	10 (59)	104 (57)	53 (39)	6 (40)	59 (39)
Other	155 (51)	16 (50)	171 (51)	71 (43)	7 (41)	78 (43)	84 (61)	9 (60)	93 (61)
Concurrent systemic therapy									
No	190 (63)	20 (63)	210 (63)	149 (90)	15 (88)	164 (90)	41 (30)	5 (33)	46 (30)
Yes	112 (37)	12 (38)	124 (37)	16 (10)	2 (12)	18 (10)	96 (70)	10 (67)	106 (70)
Fractions per week									
5	70 (23)	3 (9)	73 (22)	2 (1)	0	2 (1)	68 (50)	3 (20)	71 (47)
6	232 (77)	29 (91)	261 (78)	163 (99)	17 (100)	180 (99)	69 (50)	12 (80)	81 (53)
Neck dissection (pre-/ post-RT)									
No	201 (67)	21 (66)	222 (66)	78 (47)	10 (59)	88 (48)	123 (90)	11 (73)	134 (88)
Yes	101 (33)	11 (34)	112 (34)	87 (53)	7 (41)	94 (52)	14 (10)	4 (27)	18 (12)
Mandibular volume (median) [cc]	75.1	78.9	75.1	75.4	88.2	76.7	74.7	76.2	75.1

Within the IMRT boost group, the difference in EUD between the patients with and without ORN was limited, varying between 1.9 Gy for EUD ($\alpha = 20$) and 2.8 Gy for EUD ($\alpha = 3$) (Figure 7.3a). In the HF-SBRT boost group, the difference was larger and increased from 5.1 Gy for EUD ($\alpha = 1$) up to 7.6 Gy for EUD ($\alpha = 12$) (Figure 7.3b). Figure 7.3d shows that Dmean was higher in the IMRT boost group (Figure 7.3d). However, this difference disappeared when the dose was calculated again with EUD ($\alpha = 8$) (Figure 7.3e). AUC values for Dmean and Dmax for the IMRT boost, the HF-SBRT boost, and all patients combined were 0.59 and 0.603, 0.676 and 0.733, and 0.628 and 0.642, respectively.

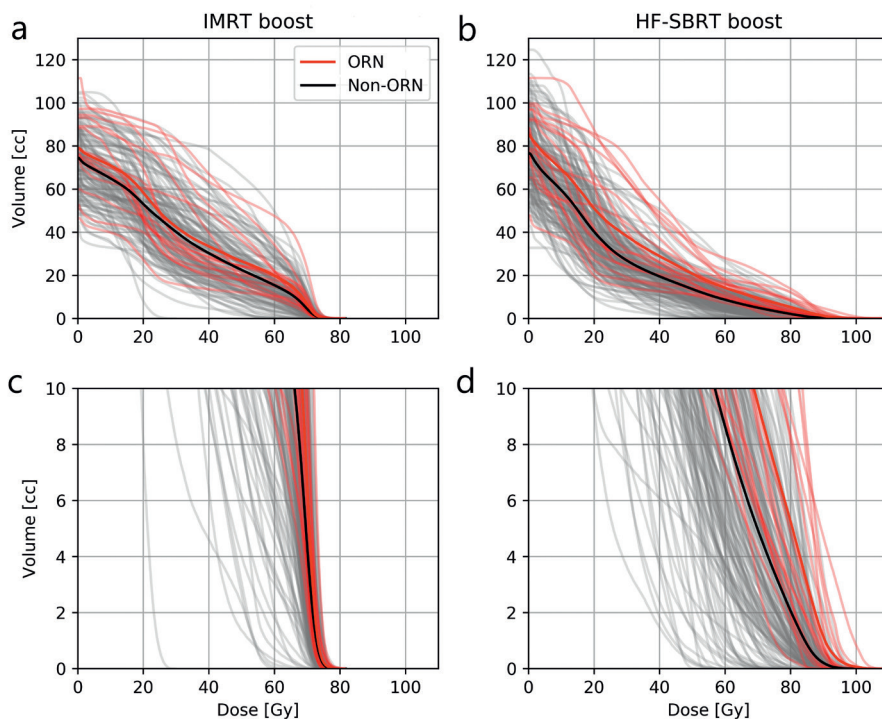


Figure 7.2. DVHs of the mandible treated with IMRT boost (a and c) and HF-SBRT boost (b and d) are shown in EQD2 for both groups. The faint lines represent the DVHs of the individual patients and the bold lines represent the DVHs averaged over the population. Figures c and d display the zoomed-in DVHs. Abbreviations: DVH = Dose-volume histogram; IMRT = Intensity Modulated Radiation Therapy; HF-SBRT = Hypofractionated Stereotactic Body Radiation Therapy; ORN = Osteoradionecrosis.

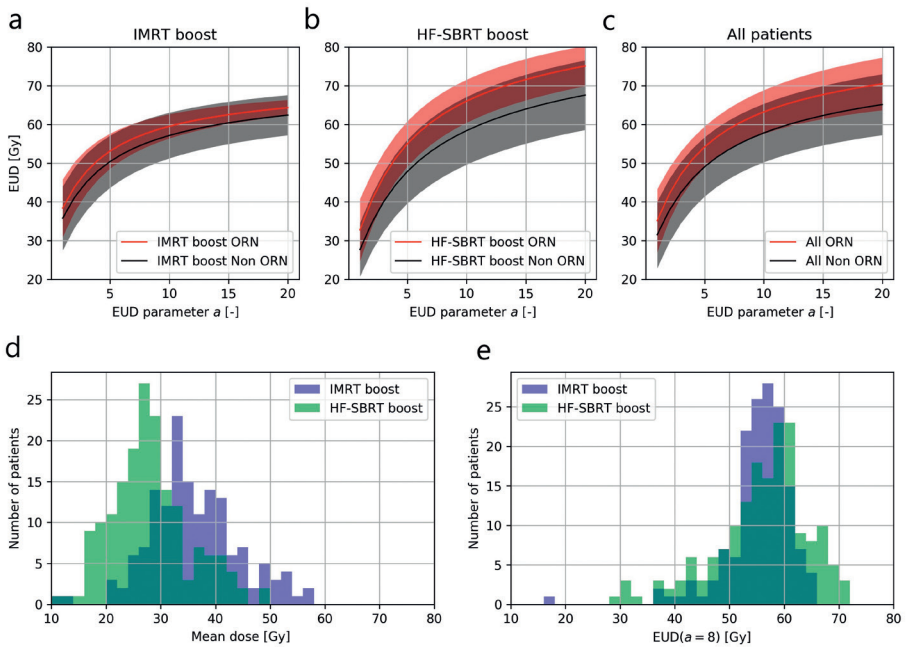


Figure 7.3. EUD for different a values for IMRT boost (a), HF-SBRT boost (b), and all patients combined (c) with and without ORN. All doses are shown in EQD2. The solid lines represent the averages over the patients with and without ORN. The shaded regions represent ± 1 standard deviation. There is a clear difference in mean dose for the IMRT boost and HF-SBRT boost (d). This difference disappears after converting the dose to EUD ($a=8$) (e). Abbreviations: EUD = Equivalent Uniform Doses; IMRT = Intensity Modulated Radiation Therapy, HF-SBRT = Hypofractionated Stereotactic Body Radiation Therapy, ORN = Osteoradionecrosis.

For the HF-SBRT boost group, the EUD (Figure 7.4a) and VGy (Figure 7.4b) dose metrics discriminated ORN from non-ORN patients better than for the IMRT boost group, reflected by higher AUC values. For the combined group, the highest AUC of 0.71 was observed for an EUD ($a=8$), and therefore EUD ($a=8$) was used for the statistical analysis. The correlations between EUD with different a parameter and VGy parameters were indeed substantial (Supporting Information Figure 7.S1).

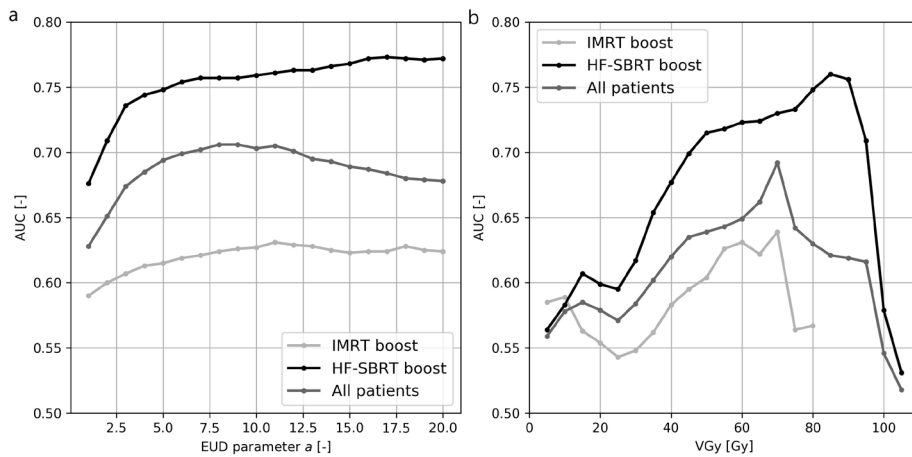


Figure 7.4. The AUC of the EUD as function of the α parameter (a) and the AUC for different VGy values (b). AUC values for Dmean and Dmax for the IMRT boost, the HF-SBRT boost, and all patients combined were 0.59 and 0.603, 0.676 and 0.733, and 0.628 and 0.642, respectively. Abbreviations: AUC = Area under the ROC curve; EUD = Equivalent Uniform Doses; VGy = volume receiving x Gray; IMRT = Intensity Modulated Radiation Therapy; HF-SBRT = Hypofractionated Stereotactic Body Radiation Therapy.

In a univariable analysis of the total patient group, pre-RT extractions (SHR = 2.69; $p = 0.005$), mandibular volume in cc (SHR = 1.03; $p = 0.018$), and EUD ($\alpha = 8$) (SHR = 1.12; $p < 0.001$) were strongly associated with ORN. Pre-RT extractions (SHR = 2.34; $p = 0.012$), mandibular volume (SHR = 1.04; $p = 0.003$), and EUD ($\alpha = 8$) (SHR = 1.14; $p < 0.010$) retained their significance in the multivariable model (Table 7.2).

Table 7.2. Risk factors for grade ≥ 2 ORN according to CTCAE v5.0. * = restricted to patients with extractions. Abbreviations: SHR = subhazard ratio; CI = Confidence Interval; WHO = World Health Organization; HPV = human papillomavirus; HF-SBRT = Stereotactic Body Radiation Therapy; IMRT = Intensity Modulated Radiation Therapy; ND = neck dissection; Dmean = mean dose; Dmax = maximum dose; V70 = volume receiving 70 Gray. Please note that for the dose parameters only the EUD was incorporated in the multivariable model as it best discriminated patients with and without ORN (highest AUG).

Total group	Univariable			Multivariable		
	SHR	95% CI	P	SHR	95% CI	P
Sex male vs. female	0.97	0.47 – 2.02	0.944			
WHO 1-2 vs. 0	0.92	0.45 – 1.88	0.820			
Alcohol current vs. no/prev	1.34	0.58 – 3.07	0.495			
Smoking current vs. no/prev	1.72	0.83 – 3.58	0.147			
Pre-RT extractions yes vs. no	2.69	1.34 – 5.40	0.005	2.34	1.20 – 4.52	0.012
Interval extractions – RT*	0.97	0.93 – 1.01	0.131			
T stage T3-4 vs. T1-2	1.10	0.54 – 2.21	0.796			
N stage N2-3 vs. N0-1	0.98	0.49 – 1.96	0.949			
Tumor subsite tonsil vs. other	1.04	0.52 – 2.08	0.912			
Systemic therapy yes vs. no	0.94	0.46 – 1.92	0.866			
RT boost HF-SBRT vs. IMRT	1.03	0.52 – 2.08	0.924			
Accelerated RT yes vs. no	2.74	0.83 – 9.04	0.099			
Neck dissection yes vs. no	1.11	0.54 – 2.31	0.774			
Mandibular volume [cc]	1.03	1.00 – 1.05	0.018	1.04	1.01 – 1.06	0.003
Mandible Dmean [Gy]	1.03	1.00 – 1.07	0.056			
Mandible Dmax [Gy]	1.13	1.04 – 1.22	0.003			
Mandible V70 Gy [cc]	1.16	1.10 – 1.22	< 0.001			
Parotid Dmean (median) [Gy]	1.00	0.97 – 1.04	0.944			
EUD $\alpha = 8$ [Gy]	1.12	1.06 – 1.18	< 0.001	1.14	1.08 – 1.20	< 0.001

7.4 Discussion

Osteoradionecrosis is a severe late complication after head and neck radiotherapy, which can have a huge impact on quality of life. The incidence and prognostic factors of ORN have mainly been analyzed in conventional fractionation schemes [55, 189, 191-195, 197, 198]. With the renewed and increasing interest in hypofractionation and the use of more heterogeneous dose distributions in head and neck radiotherapy [16, 199-201], there is a need for risk estimation models that apply also to other than conventional fractionation schemes and uniform dose distributions. Therefore, in this study patients treated with considerably different fractionation schemes and dose distributions (conventional fractionation vs. HF-SBRT) were included to explore a single model that would be appropriate for a range of fractionation schemes and dose distributions. The conversion of different fractionation schemes to 2 Gy fractions is even more important because of the low α/β for the mandible.

We used two classical established dose models. Next to the linear quadratic model, we also applied the concept of EUD to account for differences in DVH shapes. Previous studies showed that (partially) EUD-based planning provides the possibility of improved sparing of organs at risk [206-210], and in our study indeed EUD discriminated better than conventional DVH metrics between patients with and without ORN. Furthermore, EUD also helped in better understanding the volume-effect and the impact of fraction size (Figure 7.3d and 7.3e). One could expect a lower incidence of ORN in our HFSBRT boost group, because of smaller tumors treated compared to the IMRT group (86% T1-T2 vs. 30% T1-T2). However, there was no difference in the incidence rates of ORN between the two groups (9.3% vs. 9.9%), likely because of the volume effect that is incorporated in the EUD ($\alpha = 8$) (Figure 7.3d and 7.3e).

The final model included EUD with $\alpha = 8$, pre-RT extractions, and mandibular volume as significant independent risk factors of ORN.

Prior studies with conventional fractionation schemes identified the mandible receiving 50 to 70 Gy (V50-V70Gy) as the most important risk factor for ORN [55, 189, 191-195, 197, 198]. Also in the current study, it was found that in the IMRT-only group, V70Gy discriminated best between patients with and without ORN (based on AUC). Interestingly, in the HF-SBRT-only group, higher VGy values (V80Gy and V90Gy) had higher AUCs than V70Gy. This suggests that 70 Gy by itself is not the most appropriate threshold, but follows mainly from a lack of higher doses in IMRT-only cohorts, for which the prescribed dose was typically 70 Gy. The fact that V70Gy would be suboptimal for ORN prediction in a HF-SBRT dataset, demonstrates that a risk model cannot be applied

simply to dose distributions that are different from those on which the model was created. This highlights the need for more general risk models that can be applied independently to the dose distribution. Deriving such general risk models requires the inclusion of patients with variations in dose distributions and fractionation regimes, for which the current study could be considered as the first step.

For the IMRT boost group, the observed dose effect relation was less strong compared to the HF-SBRT (Figure 7.4), as suggested by the lower AUC values for both the EUD and VGy values. This might be explained by the fact that patients in the IMRT boost group had more homogeneous dose distributions, whereas in the HF-SBRT boost group the tails of the DVHs varied considerably between patients (Figure 7.2b and 7.2d).

Many (inter)national guidelines recommend pre-RT extractions in case of dental disease (e.g., periodontitis or caries profunda), because patients with extractions post-RT due to dental disease are at risk of compromised wound healing and predisposition to ORN. However, numerous studies, including this one, suggest that at the same time patients with pre-RT extractions are more likely to develop ORN [54, 55, 194, 211, 212]. If this is caused by the pre-RT extractions itself or indirectly by the existing dental disease needs to be elucidated. In case the pre-RT extractions would be an independent risk factor and not the dental disease, then a more conservative approach regarding pre-RT extractions would be warranted.

An unexpected risk factor for ORN was a larger total mandibular volume. A possible, speculative explanation might be that a larger mandibular volume reflected a (recent) dentulous situation, putting this patient group at risk for extractions post-RT, that may be in turn related to the development of ORN. Since the present study focused on pre-RT risk factors only, post-RT extractions were not considered.

A recently published study showed that differences in radiosensitivity may exist within the mandible depending on the location of dental extractions [213]. The current study, that uses EUD, did not account for these differences, but we would recommend to account for differences in radio-sensitivity within the organ in future studies.

At present, hypofractionation in head and neck cancer is mainly used in the recurrent and palliative setting, but in these situations, ORN is typically poorly documented [214–216]. Literature on ORN after radiotherapy with high fraction doses in a curative setting is even more scarce [217–221]. Eisbruch et al. reported

three patients with ORN in a cohort of 69 patients treated with a moderate hypofractionated regimen (2.2 Gy fractions), and suggested that hotspots might have provoked ORN [217]. Another paper by Karam et al. reported one patient who developed ORN grade 3 in a group of ten patients with salivary gland tumors [219].

This study has some limitations that need to be addressed. First, the study was explorative in nature and the model should be validated in an external dataset before being applied clinically. Second, we explored a general model for ORN to be applied to different fractionation schemes and types of dose distributions. Although we included two rather different schemes consisting of conventional IMRT and HF-SBRT, it is not clear whether the model could be extrapolated to other fractionation schemes.

Third, our analysis showed that an EUD ($\alpha = 8$) best discriminated between patients developing ORN and non-ORN. Since the difference in AUC for slightly lower and higher α values was limited (Figure 7.4), and high correlations were observed between different EUD values (Supporting Information Figure 7.S1), slightly different choices than $\alpha = 8$ would have likely led to similar results.

7.5 Conclusion

Our study demonstrates that risk models for ORN based on conventional DVH parameters cannot be applied directly to HF-SBRT fractionation schemes and dose distributions. We proposed here a single model for ORN that accounts for variations in fractionation and heterogeneity of the dose distributions. It was developed based on patients with either conventional or HF-SBRT treatments, and contains EUD ($\alpha = 8$), teeth extraction, and mandibular volume as factors significantly associated with ORN.

7.6 Supporting information

Table 7.S1. Classification of ORN by CTCAE v5.0. Abbreviations: ORN = osteoradionecrosis; CTCAE = Common Toxicity Criteria of Adverse Events.

CTCAE v5.0	
I	asymptomatic; clinical or diagnostic observations only; intervention not indicated
II	symptomatic; medical intervention indicated (e.g., topical agents); limiting instrumental ADL
III	severe symptoms; limiting self-care ADL; elective operative intervention indicated
IV	life-threatening consequences; urgent intervention indicated
V	death

Table 7.S2. Dose-volume characteristics for the patients with and without ORN per boost type. Abbreviations: ORN = osteoradionecrosis; IMRT = Intensity Modulated Radiation Therapy; HF-SBRT = Hypofractionated Stereotactic Body Radiation Therapy; Dmean = mean dose; Dmax = maximum dose; Gy = Gray; V30Gy = volume receiving 30 Gray; V70Gy = volume receiving 70 Gray; IQR = InterQuartile Range.

N = 325	HF-SBRT boost (N = 173)			IMRT boost (N = 152)		
	Non ORN	ORN	Total	Non ORN	ORN	Total
Mandible Dmean (median [IQR]) [Gy]	28.7 [25.5-32.6]	31.9 [27.2-40.9]	28.7 [25.7-33.1]	38.4 [33.7-43.5]	39.0 [36.1-45.1]	38.5 [33.8-43.8]
Mandible Dmax (median [IQR]) [Gy]	80.8 [78.2-83.0]	83.6 [81.2-85.8]	80.9 [78.3-83.3]	75.6 [74.7-76.4]	75.9 [75.0-77.5]	75.6 [74.8-76.5]
Mandible V30Gy (median [IQR]) [cc]	28.1 [21.2-36.3]	33.6 [25.9-63.4]	28.3 [21.3-36.6]	44.8 [35.3-56.3]	43.3 [37.4-73.9]	44.5 [35.3-56.6]
Mandible V70Gy (median [IQR]) [cc]	1.8 [0.7-3.6]	4.6 [3.2-9.9]	2.1 [0.7-4.4]	3.3 [1.5-5.9]	5.0 [3.5-8.8]	3.5 [1.7-6.2]
Parotids Dmean (median [IQR]) [Gy]	21.8 [17.2-26.4]	22.5 [18.0-25.7]	21.8 [17.3-26.4]	31.4 [26.2-38.2]	33.5 [28.1-39.5]	31.9 [26.2-38.6]

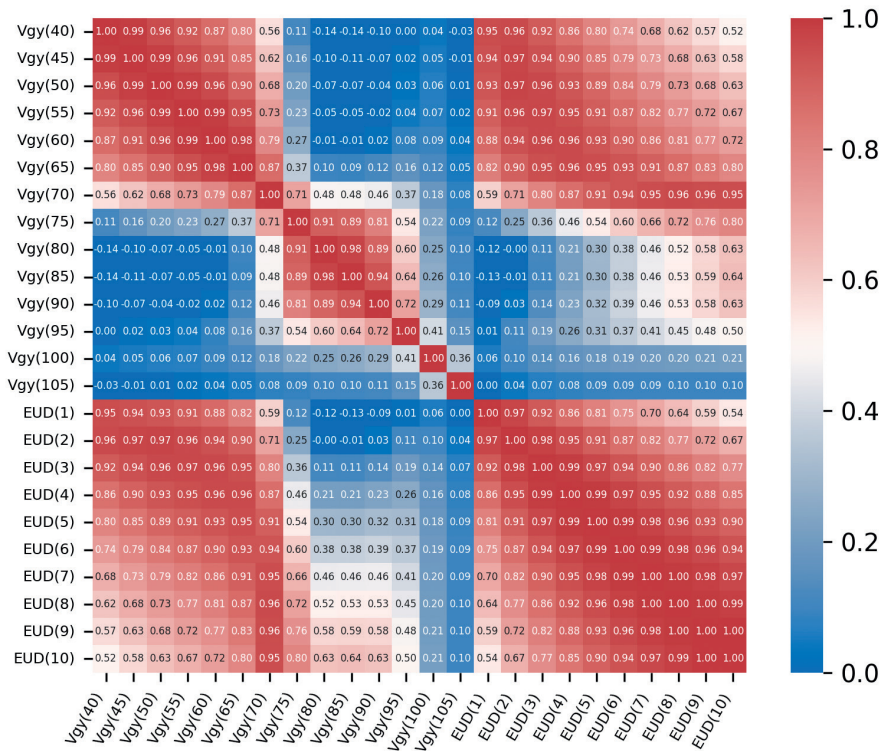


Figure 7.S1. Correlation plot between the different EUD and VGy parameters. Abbreviations: EUD = equivalent uniform dose; VGy(x) = volume receiving x Gray.

8

Development of a local dose-response relationship for osteoradionecrosis within the mandible

Nienke D. Sijtsema, Gerda M. Verduijn,
Kazem Nasserinejad, Yvette van Norden,
Hetty Mast, Aad van der Lugt,
Mischa S. Hoogeman, Steven F. Petit

This chapter is based on Sijtsema ND, Verduijn GM, Nasserinejad K, van Norden Y, Mast H, van der Lugt A, et al. Development of a local dose-response relationship for osteoradionecrosis within the mandible. Radiother Oncol. 2023;186:109736.

Abstract

Purpose

Osteoradionecrosis (ORN) of the mandible is a severe complication following radiotherapy of the head and neck, but not all regions of the mandible may be equally at risk. Therefore, our goal was to explore a local dose-response relationship for subregions of the mandible.

Materials and methods

All oropharyngeal cancer patients treated at our hospital between 2009 and 2016 were reviewed. Follow-up was cut off at 3 years. For patients that developed ORN, the ORN volume was delineated on the planning CT. Each mandible was divided into 16 volumes of interest (VOIs) based on the location of the dental elements and the presence of ORN in each was scored. Generalized estimating equations were used to build a model for the probability of developing ORN in an element VOI.

Results

Of the 219 included patients, 22 developed ORN in 89 element VOIs. Mean dose to the element VOI (odds ratio (OR)=1.05 per Gy, 95% confidence interval (CI): (1.04,1.07)), pre-radiotherapy extractions of an element ipsilateral to the element of interest (OR=2.81, 95% CI: (1.12,7.05)), and smoking at the start of radiotherapy (OR=3.37, 95% CI: (1.29,8.78)) were significantly associated with an increased probability of ORN in the VOI.

Conclusion

The developed dose-response model indicates that the probability of ORN varies within the mandible and strongly depends on the local dose, the location of extractions, and smoking.

8.1 Introduction

Osteoradionecrosis (ORN) of the mandible is a severe late complication after radiotherapy in the head and neck region. It can cause pain, oral skin fistulae, and even pathological fracture of the mandible requiring extensive reconstruction [186]. Although the incidence of ORN has declined in the era of intensity-modulated radiotherapy (IMRT), incidence rates varying between 4–10% are still reported [54, 55, 189–191]. Different studies have associated a variety of risk factors with the onset of ORN, including tumor location, size and stage, the presence of bone invasion, dose volume histogram (DVH) parameters of the entire mandible, radiotherapy technique, irradiated volume, smoking, alcohol use, oral hygiene, pre- and post-radiotherapy dental extractions, and several comorbidities [54, 55, 189–191, 193, 195].

A high dose to the mandible is related to an increased risk of ORN. However, there are indications that not all regions of the mandible are equally prone to develop ORN. For instance, ORN is most commonly observed in premolar, molar, and retro-molar areas of the mandible, but hardly in the incisor and canine area of the mandible [222]. Moreover, trauma to the mandible (i.e., due to teeth extractions) may affect the radiosensitivity of the mandible in the vicinity of the trauma. Indeed, recent reviews investigating the relationship between teeth extractions and ORN show that in some cases their locations are linked, although often the relationship is unclear [223, 224].

If differences in radiosensitivity indeed exist *within* the mandible and could be identified, selective sparing of the more sensitive regions of the mandible could reduce the likelihood of ORN. However, to the best of our knowledge, studies investigating local differences in radiosensitivity within the mandible are currently lacking. Therefore, the goal of the current study was to explore a local dose-response relationship taking into account the location of ORN, the dose deposited locally to the ORN site, and the vicinity of teeth extractions.

8.2 Materials and methods

8.2.1 Patients

The records of all patients with oropharyngeal squamous cell carcinoma who were treated with curative (chemo)radiotherapy at our department between January 2009 and May 2016 and survived at least one year were reviewed. Exclusion criteria were as follows: diagnosis of another primary tumor within six months, previous oropharyngeal cancer, previous head and neck radiotherapy, tumor progression within six months, unavailability of dosimetric data or radiological data of ORN location, follow up of less than three years except

for patients that developed ORN within three years. Patients not followed-up in our hospital were not actively approached. The study was approved by the institutional review board (protocol EMC17404).

8.2.2 Treatment

Patients were referred to either one of two treatment schedules in line with the standard clinical protocol. Patients with large cT3–T4N0–2M0 and any cT1–4N3M0 tumors received IMRT of 70 Gy IMRT (35 fractions of 2 Gy) to the primary tumor and regions containing pathological neck nodes, and 46 Gy to the elective neck regions, with a sequential or simultaneously integrated boost. At least 98% of the PTV should be covered by 95% of the prescription dose ($V_{95} > 98\%$). Chemotherapy was added to the treatment if indicated based on TNM stage (T3–4 or N+), and consisted of cisplatin (100 mg/m² on days 1, 22, and 43 of treatment) or cetuximab (400 mg/m² initial dose, followed by a weekly dose of 250 mg/m²). The overall treatment time of this group was 6 weeks in case of accelerated treatment (6 fractions/week) and 7 weeks in case of non-accelerated treatment (5 fractions/week).

Patients with cT1–smallT3N0–2cM0 tumors were treated with 46 Gy IMRT (23 fractions of 2 Gy) to the primary tumor and elective neck node regions, followed by a stereotactic body radiotherapy (SBRT) boost of 16.5 Gy to the primary tumor (3 fractions 5.5 Gy) delivered by the Cyberknife system (Accuray Inc., Sunnyvale, CA, USA). The SBRT boost dose of 16.5 Gy was prescribed to the 80% isodose line, where the maximum dose was set to 100%. The overall treatment time for this group was 5 weeks. In the SBRT boost group, patients with an N+ neck underwent neck dissection after the boost.

Patients were routinely seen at the Department of Oral and Maxillofacial Surgery before the start of radiotherapy and dental extractions were performed according to national guidelines of the Dutch Association of Oral and Maxillofacial Surgeons [225]. During radiotherapy treatment, dentulous patients were seen weekly by a dental hygienist, and received fluoride treatment. Patients were followed by the head and neck multidisciplinary team. Follow-up visits were planned every 2 months for the first year following radiotherapy. Starting from the second year, the frequency gradually decreased to every 6 months for a minimum of 5 years. In case of suspicion or diagnosis of ORN, patients were referred to the ORN outpatient clinic at the Department of Oral and Maxillofacial Surgery.

8.2.3 Scoring and delineation of ORN

ORN was defined as clinically exposed bone following radiotherapy, without evidence of recurrent or residual tumor and no signs of healing for at least

three months. ORN as late toxicity (> 90 days after completion of radiotherapy) was scored according to CTCAE v5.0 by an experienced radiation oncologist. Predictors were scored independently, before knowing whether a patient had ORN, and were therefore blinded for outcome. Follow-up was cut off at three years. For patients with ORN grade 2 or higher within three years after treatment, regions of mandibular bone affected by ORN (ORN volumes of interest (VOIs)) were delineated by a radiation oncologist and an oral and maxillofacial surgeon (both experienced in ORN) on the planning CT, based on clinical and radiological examination of the mandible.

8.2.4 Scaling and accumulation of dose distributions to the mandible

All dose distributions were converted voxel-wise to the equivalent dose of 2 Gy (EQD2Gy) based on $\alpha/\beta=0.85$ Gy for bone toxicity [202]. Patients in the IMRT boost group had one (simultaneous integrated boost) or two (sequential boost) treatment plan(s) planned on the same planning CT, that were accumulated for the purpose of the study.

Patients in the SBRT boost group had at least two planning CT scans and treatment plans: one corresponding to the IMRT and one corresponding to the SBRT part of the treatment. Therefore, dose accumulation required registration between the CT scans. The following procedure was used. First, the mandible was automatically delineated in both scans using ADMIRE 3.7.7 (Elekta AB, Stockholm, Sweden) with employment of STAPLE [226]. The delineations were assessed qualitatively before being used in the registration procedure. Subsequently, the rigid registration (rotation and translation) was performed based on the delineations of the mandible. After registration, the EQD2Gy SBRT and IMRT dose distributions to the mandible were accumulated.

For patients that had adaptive treatment plans, i.e. patients that showed considerable anatomical changes during treatment requiring a new treatment plan, dose was accumulated according to the same methods used for the SBRT boost group.

8.2.5 Transforming all dose distributions and delineations to one reference patient

To investigate the effect of anatomical location on the risk of ORN, the accumulated dose distributions and ORN delineations of all patients were projected on top of a reference patient using deformable image registration (ADMIRE 3.7.7 (Elekta AB, Stockholm, Sweden)). The accuracy of the deformable registration was assessed by comparing the mandible delineation of the reference patient with the transformed mandible delineations.

Next, the mandible of the reference patient was divided into different subvolumes based on the location of the dental elements. For that purpose, all dental elements of the mandible were delineated on a patient with full dentition (including third molars and without severe crowding), because it was not feasible to accurately manually delineate these subvolumes in patients with partial or no dentition. To cover as much of the mandible as possible, the delineations were extended in caudal direction up to the caudal edge of the mandible. These regions were referred to as element VOIs and were numbered according to the WHO ISO dental notation system (left: 3.1 – 3.8, right: 4.1 – 4.8), where .1 refers to the central incisor and .8 to the third molar. The element VOIs were subsequently transformed to the reference patient. ORN at an element was scored if there was any overlap between the ORN VOI and the element VOI.

8.2.6 Statistical analysis

The probability of ORN at three years was modeled using a generalized estimating equations (GEE) model. GEE is an extension of generalized linear models for the analysis of clustered data, such as multiple measurements per patient, in our case multiple VOIs per patient [227, 228]. GEE is based on a quasi-likelihood function and provides population-averaged estimates of the model coefficients. A first-order autoregressive (AR1) working correlation matrix was chosen to capture the inherent correlation between dental elements within the subjects. The AR1 assumes the correlation between any two elements is equal to ρ for adjacent elements, ρ^2 for elements that are separated by one element, and so on. This structure is suitable for the purpose of the study since it acknowledges that if a certain element contains ORN, neighboring and nearby elements may have an increased risk of having ORN too. To tackle the potential misspecification of the covariance structure, the robust (sandwich) estimator [227] was used to produce unbiased standard error estimators for regression coefficients.

For the selection of patient-level variables, the variables age, gender, WHO performance status (0 vs 1 and 2), alcohol use (yes/no), smoking at the start of therapy (yes/no), pretreatment teeth extractions (yes/no), T stage (1 and 2 vs 3 and 4), HPV status (positive/negative), treatment acceleration (5 vs 6 fractions per week), boost type (IMRT vs SBRT), concurrent systemic therapy, neck dissection, and tumor subsite (tonsil vs other) were first separately tested on a patient level in univariate logistic regressions.

All patient-level variables with a $p < 0.05$ in the univariate analysis were taken into account in the GEE model, which accounted for ORN on the element level. The following element VOI variables were taken into account in the GEE

model: (1) The mean dose to the element VOI, and (2) whether the element was a molar (yes/no), or a (3) premolar (yes/no) to account for the anatomical location. Teeth extractions (in case univariable significant) were taken into account locally through the following three variables: (1) whether the element of the element VOI was extracted pretreatment (yes/no), (2) any pretreatment extractions ipsilateral to the element VOI including extractions of the element of interest itself (yes/no) and (3) any pretreatment extractions on the contralateral side of the mandible (yes/no).

Variable selection for the final GEE model was done using five-fold selection. To that end, the group of ORN and non-ORN patients were first separately divided into five folds. Next, the folds of the ORN and non-ORN patients were combined. In this way, the original distribution of ORN and non-ORN patients was maintained in each of the folds. Next, five models were built, each on four out of the five folds, leaving out a different fold for each model (leave one out approach), and backward selection was performed for variable selection for each of the models. If a variable was significant in the majority (at least three) of the five models after backward selection, the variable was included in the final model. In the final step, the final model coefficients of the selected variables were determined by fitting the GEE model to all data. To demonstrate the size of the effects, the model coefficients and odds ratios (ORs) are presented along their confidence intervals.

To study the robustness of the model stratified cluster bootstrapping (500 times) was performed. In stratified cluster bootstrapping the percentage of ORN and non-ORN patients included in each bootstrapping sample was the same as the percentage of ORN and non-ORN patients in the original data. The bootstrapping confidence intervals of the coefficients are reported. The receiver-operator characteristic (ROC) curve was calculated and reported with its corresponding confidence interval (based on 2000 bootstrap samples) to assess discrimination of the model. A calibration plot was made to assess the calibration of the model. All statistical analyses were performed using R Statistical Software (version 4.1.2; R Foundation for Statistical Computing, Vienna, Austria) using *geepack* (version 1.3-2), *pROC* (version 1.18.0), and *rms* (version 6.2-0) packages. A p -value < 0.05 was considered statistically significant.

8.3 Results

Patient and treatment characteristics are summarized in Table 8.1, as well as the results from the patient-level univariate analysis. In total, 334 patients were reviewed, of which 219 were selected after applying the exclusion and

follow-up criteria. One ORN patient was excluded due to unavailability of radiological data on the site of ORN. 120 patients (54.8%) were treated with the SBRT boost protocol and 99 patients (45.2%) with the IMRT boost protocol. For these patients, there were no missing data for all predictors and outcomes. Twenty-two patients developed ORN CTCAE v5.0 grade 2 or higher within three years (11 grade 2, 11 grade 3). One patient developed bilateral ORN. The median volume of the ORN VOIs was 6.8 cc (range 0.6–43.2 cc). On average the ORN VOI covered 10.9% of the mandible (range 1.1–44.5%). The average symmetric distance between mandible delineations after registration of the IMRT CT and the SBRT CT was on average 1.0 mm (range 0.6–2.0 mm), indicating sufficiently accurate registrations.

Table 8.1. Patients and treatment characteristics. ORN was scored according to CTCAE v5.0. ORN - osteoradionecrosis; WHO - World Health Organization; HPV - human papilloma virus; SBRT - Stereotactic Body RadioTherapy; IMRT - Intensity Modulated Radiotherapy.

Characteristics	All patients	Non-ORN	ORN (≥grade 2)	p-value
N	219	197	22	
Age (mean; SD)	61;9	61;9	64;7	0.47
Gender				0.90
Male	148 (54.8%)	133 (67.5%)	15 (68.2%)	
Female	71 (45.2%)	64 (32.5%)	7 (31.8%)	
WHO performance status				0.84
0	136 (62.1%)	125 (63.5%)	11 (50.0%)	
1-2	83 (37.9%)	72 (36.5%)	11 (50.0%)	
Alcohol				0.27
No/ previous	57 (26.0%)	54 (27.4%)	1 (4.5%)	
Current	162(74.0%)	143 (71.6%)	21 (95.5%)	
Smoking				0.041*
No/ previous	116 (53.0%)	109 (55.3%)	2 (9.1%)	
Current	103 (47.0%)	88 (44.7%)	20 (90.9%)	
Teeth extraction				0.022*
No/ edentulous	139 (63.5%)	132 (67.0%)	7 (31.8%)	
Partly/ completely	80 (36.5%)	65 (33.0%)	15 (68.2%)	
Time (days) extr – RT (median)	20	20	17	
T stage				0.73
T1-T2	132 (60.3%)	121 (61.4%)	11 (50.0%)	
T3-T4	87 (39.7%)	76 (38.6%)	11 (50.0%)	

Table 8.1. *Continued.*

Characteristics	All patients	Non-ORN	ORN (\geqgrade 2)	p-value
HPV (P16) status				0.16
Negative	24 (11.0%)	20 (10.2%)	4 (18.2%)	
Positive	73 (33.3%)	69 (35.0%)	4 (18.2%)	
Unknown	122 (55.7%)	108 (54.8%)	14 (63.6%)	
Fractions per week				0.036*
5	46 (21.0%)	44 (22.3%)	2 (9.1%)	
6	173 (79.0%)	153 (77.7%)	20 (90.9%)	
RT boost				0.47
SBRT	120 (54.8%)	110 (55.8%)	10 (45.5%)	
IMRT	99 (45.2%)	87 (44.2%)	12 (54.5%)	
Concurrent systemic therapy				0.80
No	135 (61.6%)	122 (61.9%)	13 (59.1%)	
Yes	84 (38.4%)	75 (38.1%)	9 (40.9%)	
Neck dissection (pre-/ post RT)				0.74
No	145 (66.2%)	130 (66.0%)	15 (68.2%)	
Yes	74 (33.8%)	67 (34.0%)	7 (31.8%)	
Tumor subsite				0.95
Tonsil	103 (47.0%)	94 (47.7%)	9 (40.9%)	
Other	116 (53.0%)	103 (52.3%)	13 (59.1%)	

Out of 3504 dental element VOIs in 219 patients, 89 element VOIs contained ORN. For the patients with ORN, on average 4 dental element VOIs (range 2-13) were affected. ORN was most frequently found in the molars, followed by the premolars, and more often on the right side than the left side (Figure 8.1a and b). The voxel-wise mean dose to the mandible for the group with and without ORN is shown in Figure 8.1c and 8.1d, respectively, and the difference in mean dose in Figure 8.1e. The average dose on the right side of the mandible was higher for patients with ORN, compared to those without ORN, which was in line with a higher incidence of ORN on the right compared to the left side of the mandible.

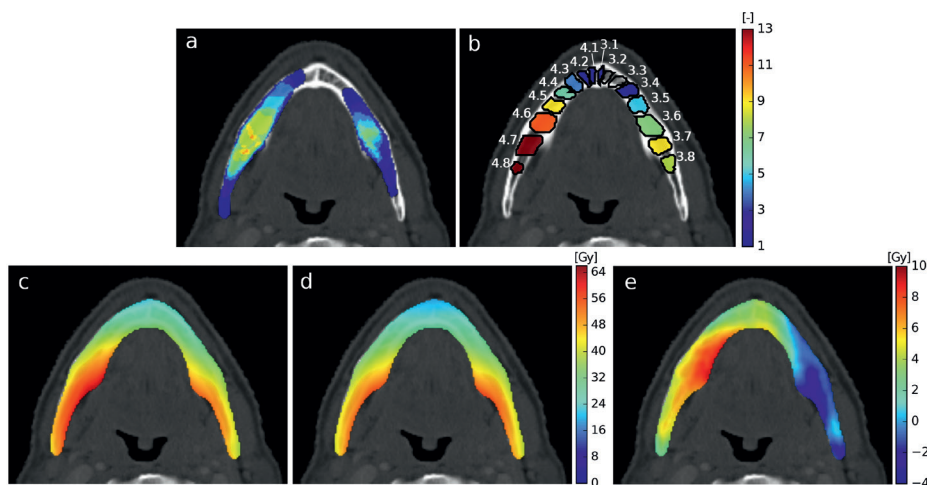


Figure 8.1. Population map of the number of voxels with ORN and average dose distributions, deformed to and projected on the reference patient. (a) Total number of ORN cases depicted voxel-wise. (b) The number of ORN cases per delineated dental element VOI, the delineations of each dental element, and their corresponding number. Note that none of the patients had ORN in elements 3.2 and 3.3. Average dose maps [EQD2Gy] of the mandible of (c) patients with ORN and (d) patients without ORN. The asymmetry of the average dose map of patients with ORN (c) is most likely due to the majority (13 out of 22) of tumors being located on the right side of the mandible in the group of patients that developed ORN. In (e) the difference in average dose (ORN – no ORN) [EQD2Gy] is shown.

From the univariate patient-level analysis, smoking at the start of radiotherapy, accelerated treatment, and teeth extractions were significant (Table 8.1). After parameter selection, the final multivariate model contained mean dose to the element VOI (OR = 1.05 per Gy, 95% CI: (1.04, 1.07)), pre-radiotherapy extractions ipsilateral to the element of interest (including extraction of the element of interest itself) (OR = 2.81, 95% CI: (1.12, 7.05)), and smoking at the start of radiotherapy (OR = 3.37, 95% CI: (1.29, 8.78)). Table 8.2 contains the details of the final model. The estimated correlation coefficient ρ was 0.61. Supporting information Figure 8.S1 shows the ROC curve of the final model. Supporting information Figure 8.S2 contains a calibration plot and several calibration measures. Please note that both the ROC curve and calibration measures were based on the same data as used for model development, and that the estimated risks were population average probabilities rather than patient individual probabilities. The corresponding population average dose-response relations are shown in Figure 8.2.

Table 8.2. Intercept and coefficients for the final model including mean dose to the element VOI, extraction of elements on the same side of the mandible (including extraction of the element of interest itself), and smoking at start RT as variables. CI – Confidence interval, EQD2Gy – equivalent dose of 2 Gy, OR – Odds ratio.

	Coefficient (95% CI)	OR (95% CI)	p-value	Bootstrapped Coefficient CI	Bootstrapped OR CI
Intercept	-7.01 (-8.04, 5.99)		< 10 ⁻³		
Mean dose to element VOI [EQD2Gy]	0.050 (0.040, 0.063)	1.05 (1.04, 1.07)	< 10 ⁻³	(0.037, 0.069)	(1.04, 1.07)
Ipsilateral extraction(s)	1.03 (0.11, 1.95)	2.81 (1.12, 7.05)	0.028	(-0.13, 2.01)	(0.88, 7.43)
Smoking at start RT	1.21 (0.25, 2.17)	3.37 (1.29, 8.78)	0.013	(0.40, 2.47)	(1.49, 11.85)

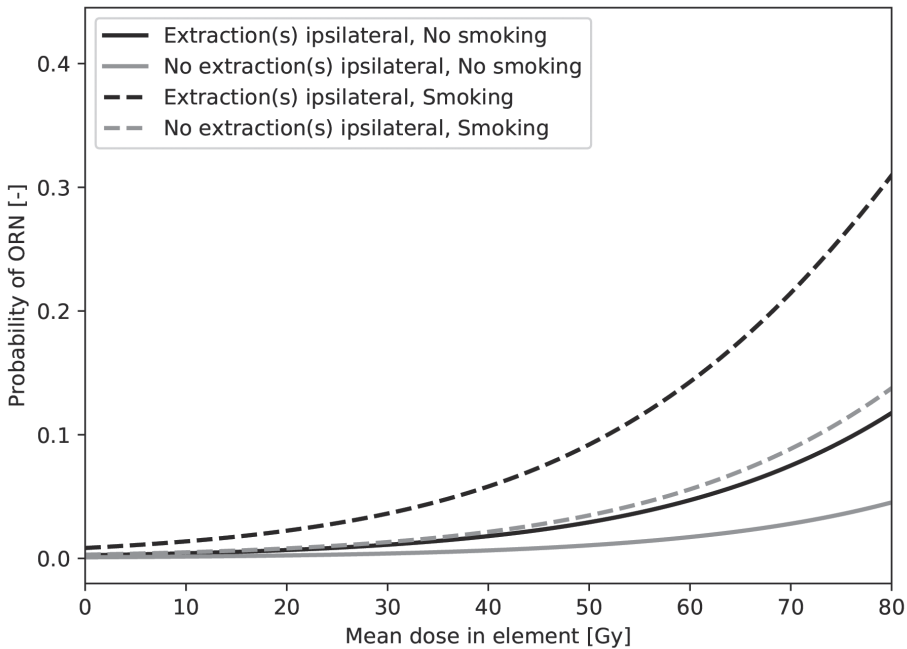


Figure 8.2. Plots of the population average probability of ORN as function of the mean dose in the element VOI [EQD2Gy]. The different curves represent situations with (black) and without (gray) extractions on the ipsilateral side of the element for non-smokers (solid) and smokers (dashed). The confidence intervals for the model coefficients can be found in Table 8.2.

8.4 Discussion

ORN is a severe complication following radiotherapy of the head and neck. ORN mainly manifests in the posterior mandible (molars and premolars) [195, 222] (supported by the current study) and teeth extractions are an important risk factor. This suggests that not all regions of the mandible are equally prone to ORN. Therefore in this study we derived a local dose-response relationship of the mandible, corresponding to the location of dental elements. We found that the dose to dental element VOIs, extractions of and/or near the element of interest, and smoking at the start of treatment were strong predictors of ORN at the element of interest.

The incidence of ORN in our cohort was similar to De Maesschalck et al. [190] who found a 10% incidence of ORN in a cohort treated with IMRT. Though, other recent studies have reported slightly lower incidence values (4.6-7.5%) [54, 55, 189, 229]. Previous studies that looked into risk factors for ORN focused on

the *patient-level* variables, and found high doses to the mandible (e.g., mean dose), the presence of teeth extractions anywhere in the mandible (yes/no), and smoking as relevant risk factors [54, 55, 189, 193, 195, 229, 230]. However, none of these studies have investigated whether the exact location of dose deposition and teeth extractions is related to the location (and presence) of ORN or not. If there would be a location effect, sparing complication-prone regions of the mandible (for instance where extractions took place) may reduce the risk of ORN. However, if there would be no effect of location, selective sparing of certain mandible regions would not be effective. In the current study, we demonstrate that in our data a local dose-effect relation in the mandible does exist and that the effect depends on whether dental elements in the vicinity are extracted and, on a patient level, whether the patient smoked or not.

We found that the anatomical location (molar vs premolar vs teeth) was not significant in our analysis. This suggests that the higher frequency of ORN observed in the molars is not due to its anatomical location, but most likely due to the dose that is often higher in the posterior mandible compared to the other mandible regions. We found ORN more frequently on the right side of the mandible, which was in line with the fact that most tumors in the group of patients that developed ORN were also located on the right side (13 out of 22) which led to on average a higher dose on the right side.

Previous studies have proposed dose constraints for the entire mandible, implicitly assuming no differences in radiosensitivity within the mandible. For instance, Owosho et al. [195] suggested a maximum dose of 60 Gy on the mandible in a group of patients that developed ORN with very few pre-radiotherapy extractions (2/44 patients). Due to tumor location, in our dataset for only 2% of the patients a maximum dose lower than 60 Gy was achieved (without setting constraints for the mandible). So, a 60 Gy dose constraint does not seem feasible for the majority of our oropharyngeal cancer patients. Mohamed et al. [189] suggested $V_{44} < 42\%$ and $V_{58} < 25\%$ as constraints for the mandible, which according to a recent paper by Lee et al. [198] was feasible in the majority of their OPC patients. While such a DVH constraint will diminish high dose areas within the mandible, the remaining high dose being delivered close to teeth extractions might still lead to an increased risk of ORN. So, a DVH constraint of the whole mandible disregards that the location of teeth extractions could play an important role in the risk of ORN.

Indeed, our results suggest that the risk of ORN does depend on the location of the extractions. Having extractions on the contralateral side did not increase the risk compared to no extractions (OR non-significant), while extractions

on the ipsilateral side or of the element of interest did increase the likelihood of developing ORN (OR=2.81, 95% CI: (1.12,7.05)). The ORs from GEE can be interpreted the same way as in a logistic regression model. The model therefore suggests that steering away high doses from areas with extractions would reduce the risk of ORN.

While GEE appropriately accounts for the correlation between dental elements of a patient, a limitation of GEE is that it allows for population average predictions only. Since estimating only population average effects is computationally easier than estimating both population average (fixed) and patient random effects, GEE is better capable of dealing with smaller datasets than, for example, mixed-effect models. An additional implication of the fact that GEE only estimates population average effects is that it does not provide a way to translate the probability of ORN for each of the elements into a probability of ORN on a patient level. The correlation between elements in one patient ($\rho=0.61$) suggests that ORN in these elements can indeed not be regarded independently, which implies that the element-level probabilities are also not additive. To be able to use these results in the clinic, for example in treatment planning, it would be necessary to fit a multi-level mixed-effect model (i.e. a model that includes both random and fixed effects) to enable patient-specific predictions followed by external validation of the model. Such a multi-level mixed-effect model would require a considerably higher number of patients and events compared to the current study with 219 patients, and is, therefore, a topic for future research. What can be concluded from the GEE model is that further (multi-center) studies into ORN should acknowledge the location of extractions, the location of ORN within the mandible and the locally deposited dose, and not simply consider the mandible as one structure as most studies have done up to this point.

This study has a couple of additional limitations that need to be addressed. First, the bootstrapping confidence intervals of the coefficients are generally slightly broader than those calculated based on the robust standard error. This was expected due to the limited data size. A clustered, stratified bootstrap was done to ensure the percentage of patients with ORN would be the same in all bootstrap samples. However, no stratification with regard to the distribution of the extractions or smoking over the bootstrap sample was deployed. Since for smoking only two patients that developed ORN were non-smokers in our dataset, it is unlikely that the bootstrap sample contained only non-smokers or only very few smokers that developed ORN. In the case of extractions, it is possible that some bootstrap samples contained none or almost no patients with extractions ipsilateral to the element of interest, since around 33% of the patients that developed ORN did not have extractions pre-radiotherapy. This

could explain why the resulting bootstrap coefficient CI included OR=1 for extractions.

Second, a follow-up cut-off was chosen instead of time-to-event analysis. This was due, to the best of our knowledge, to the unavailability of a method that combined GEE with an AR1 working correlation matrix with time-to-event analysis.

Third, in this study, we used the mean dose instead of the maximum dose to each dental element VOI as the mean dose is generally more robust than the maximum dose to deviations between planned and delivered dose. Also, any treatment adaptations were fully accounted for in the dose accumulation. However, there still may be deviations between the planned and delivered dose. Fifth, even though post-radiotherapy extractions are a known risk factor for development of ORN, they were not taken into account in the analysis, since they are generally not known at the time of treatment planning, and can therefore not be accounted for to selectively spare regions at risk.

Finally, since patients with previous head and neck radiotherapy, as well as patients with previous oropharyngeal cancer were excluded from this study, the model does not apply to these two groups of patients. Additionally, one ORN patient was excluded because no radiological data on ORN location was available, which could have led to a slight underestimation of the ORN incidence.

8.5 Conclusion

We derived a local dose-response relationship on the level of the dental elements. Elements with ipsilateral extraction(s) had 2.81 (95% CI: (1.12, 7.05)) times higher odds of developing ORN than elements without ipsilateral extractions. Smoking at the start of radiotherapy (OR=3.37, 95% CI: (1.29, 8.78)), and mean dose to the element (OR=1.05 per Gy, 95% CI: (1.04, 1.07)) also significantly increased the odds of developing ORN. Our results indicate that regions of the mandible with pretreatment extractions are locally more susceptible to ORN than regions further away from the extractions. In future work, the local variation in radio-sensitivity should be taken into account in a larger dataset allowing for modeling of both fixed (population average) and patient random effects. Once independently validated, the dose-response relations could help to selectively spare sensitive regions of the mandible to reduce the risk of ORN for patients with pre-radiotherapy extractions.

8.6 Supporting information

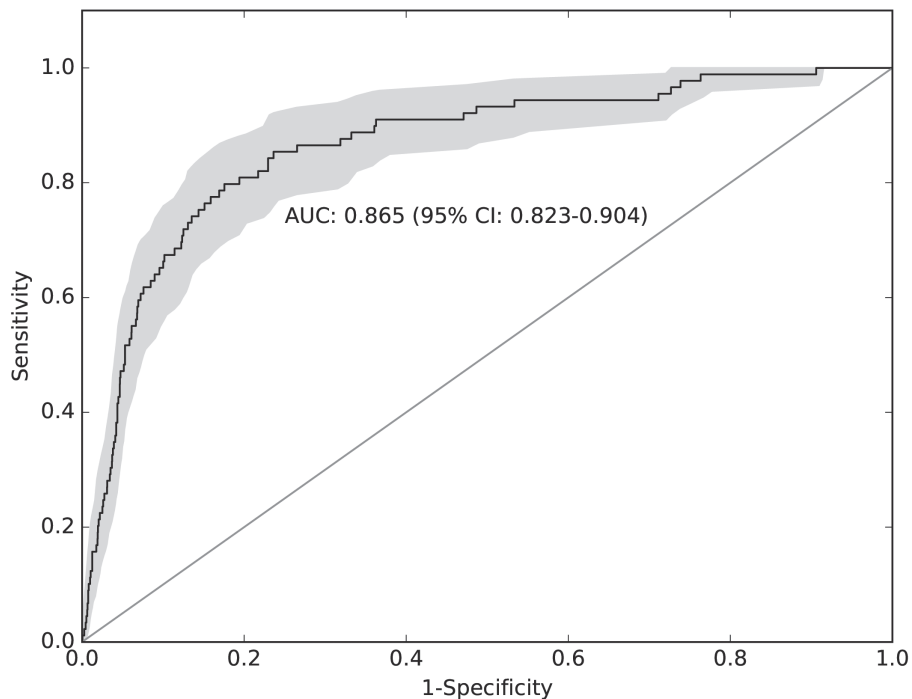


Figure 8.S1. ROC curve of the final model shown in black with the corresponding 95% confidence interval in gray. The AUC was 0.87 (95% CI: 0.82–0.90). Please note that the ROC curve is based on the same data as used for model development, and that the estimated risks were population average risks rather than patient individual risks.

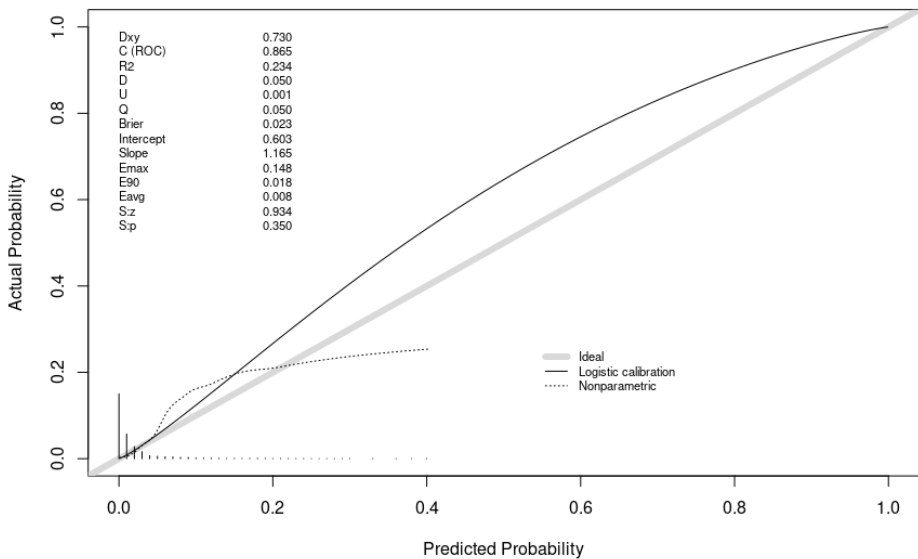


Figure 8.S2. Calibration plot and measures for the presented GEE model. Please note that the plot and measures were based on the same data as used for model development, and that the estimated risks were population average probabilities rather than patient individual probabilities. The ticks on the x-axis represent the distribution of predicted probabilities. The logistic calibration line and the non-parametric line follow the ideal calibration line only at low probabilities, which is likely due to the scarcity of data at higher predicted probabilities. Dxy, Somer's Dxy; C (ROC), AUC; R2, R-squared; D, discrimination index; U, Unreliability index; Q, quality index; Brier, brier score; Intercept, calibration intercept; Slope, calibration slope; E_{max}, maximum absolute difference in predicted and loess-calibrated probabilities; E₉₀, 0.9 quantile of absolute difference in predicted and loess-calibrated probabilities; E_{avg}, average absolute difference in predicted and loess-calibrated probabilities; S:z and S:p, Spiegelhalter's Z-test Z and p-value respectively.

9

Discussion

9.1 Overview

This thesis focused on the development of imaging-based biomarkers and models that provide insight into the prediction of tumor response and treatment-related toxicity. By using these response predictors in the future, we aim to optimize treatment planning and decision-making for patients individually, ultimately leading to improved patient outcomes.

As described in the COMPLETE study protocol presented in **Chapter 5**, tumor response is multi-factorial and heterogeneous among patients. Obtaining a complete picture of tumor response to treatment likely requires techniques describing different aspects of tumor response. The first part of this thesis focused on the prediction of tumor response in the macroscopic domain using Non-Gaussian Intravoxel Incoherent Motion diffusion-weighted imaging (NG-IVIM DWI) and multi-delay pseudo-continuous arterial spin labeling (pCASL) MRI. On the other hand, a key aspect of treatment personalization is establishing dose objectives for organs at risk (OARs) based on patient-specific risk of toxicity. Therefore, the second part of this thesis focuses on improving our understanding of treatment-related toxicity, in particular for osteoradionecrosis.

After thorough clinical validation, the response predictors presented in this thesis could be used to select patients for certain personalization strategies, for example: (1) more dose to tumors that need more dose, (2) less dose to tumors that need less dose, (3) higher biological dose, (4) more dose to patients that can tolerate more dose, (5) selective sparing of complication prone regions. Figure 9.1, which is a repeat of Figure 1.2, outlines how each of the works in this thesis fits with the mentioned personalization approaches.

The work presented in this thesis, as well as the encountered challenges, remaining considerations and opportunities, and future directions will be discussed in this chapter.

9.2 MRI-based imaging biomarkers

MRI is a very versatile tool providing both anatomical and biological information in one comprehensive scanning protocol, allowing for obtaining key information on baseline functionality and early response of several tissues. This information could then be used to determine the optimal treatment. Many techniques have been investigated over the years, including the two techniques used in part one of this thesis: NG-IVIM DWI and multi-delay pCASL. In this thesis, we selected NG-IVIM and multi-delay pCASL, due to their (theoretical)

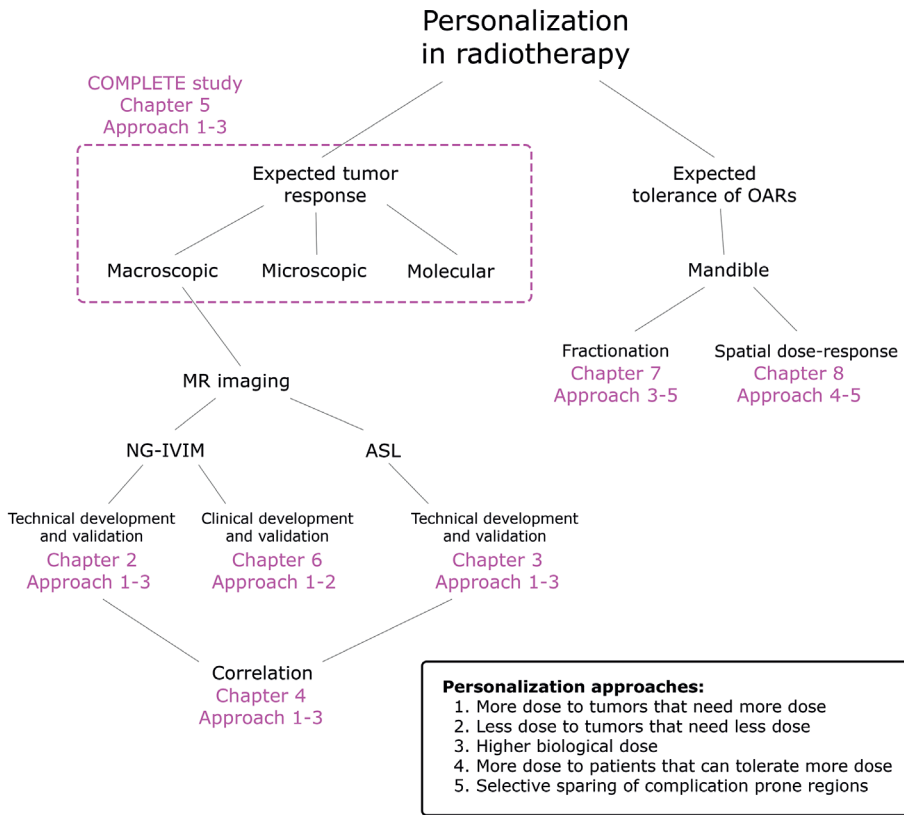


Figure 9.1. Personalization in radiotherapy depends on both the expected tumor response on a macroscopic, microscopic, and molecular level (dashed box), as well as the expected tolerance of OARs. For the prediction of tumor response, this thesis focused on implementing and optimizing non-Gaussian intravoxel incoherent motion imaging (NG-IVIM) and arterial spin labeling (ASL) for prediction of the macroscopic tumor response. Additionally, the correlation between these two techniques was investigated. For the prediction of the tolerance of OARs, we specifically focused on the mandible, and investigated the influence of fractionation and spatially varying dose on the risk of osteoradionecrosis. Note that this figure is identical to Figure 1.2.

ability to measure perfusion and diffusion, both characteristics that are typically disrupted in tumors, as well as based on available data from previous studies from other groups [56, 92].

O'Connor et al. [231] outline three domains for establishing imaging-based biomarkers: (1) discovery (2) technical and clinical validation, and (3) qualification for clinical use. The work in this thesis focused on domain two, technical and clinical validation. Domain 2 can be further separated into several

steps: (1) single-center technical development and validation, (2) single-center clinical validation, (3) translation to multi-center clinical research tool, and (4) multi-center technical and clinical validation. Specifically, the work described in the thesis for NG-IVIM focused on domain 2, steps 1 and 2 while keeping in mind step 3. The work on ASL focused on domain 2, step 1.

9.2.1 NG-IVIM DWI

Single-center technical development and validation

In **Chapter 2** we optimized b-values using a scanner-independent optimization strategy based on the Cramér-Rao lower bound, a measure for variance. This resulted in a set of 15 b-values that can be used as a starting point for standardization of DWI scan parameters across centers. This was important because the lack of standardization is a major reason large, multi-center studies are currently lacking, and to fully establish the role of DWI in the clinic multi-center studies are vital. Additionally, the optimization strategy could straightforwardly be used for other DWI models. For instance in **Chapter 4**, we used a b-value set optimized specifically for IVIM using the optimization strategy presented in **Chapter 2**.

Although DWI is one of the few biological imaging techniques that is generally available as a standard sequence, different MR scanners, especially those of different vendors, often have slightly different implementations of an imaging technique. This introduces variability among the measurements, even when using the same imaging parameters. It is important to quantify such variability to be able to compare the results from different scanners and even centers, and QA procedures will need to be established.

Also in **Chapter 2**, we looked into post-processing of NG-IVIM. Specifically, we compared different motion correction strategies applied to NG-IVIM images using elastix [78], an open source software for image registration. Although we showed that registration did not improve the precision of parameter estimation, by using open source software we could ensure that our approach could easily be adopted by others. This addresses a limitation of many processing strategies; that it is done with in-house developed solutions, since either there is no vendor solution available, or the processing provided by the vendor lacks transparency and customizability. The lack of a well-established, reliable, and generalizable processing framework inherently gives rise to a large heterogeneity in reported measures and processing techniques used. By using published pipelines and open source tools the processing steps, such as motion correction, can be incorporated in a standardized manner across centers.

The fitting of DWI signal to the IVIM or NG-IVIM model is not addressed in this thesis, but is an important step. Both the IVIM and NG-IVIM are difficult to fit due to their bi-exponential nature, and multiple solutions could yield a similar fitting result. To enable the fitting algorithm to find more robust and physiologically plausible solutions, different strategies have been employed such as the variable starting point algorithm used in this thesis in **Chapters 2, 4, and 6**, or a two-step approach where first the diffusion coefficient is estimated and fixed before the other parameters are estimated [232, 233]. Additionally, artificial intelligence could provide an alternative to conventional fitting algorithms [234]. As of yet, no consensus on the best fitting strategy has been reached. Since different fitting algorithms give rise to variability among the measurements, the fitting algorithm should also be standardized in multi-center studies.

In **Chapter 2** we also addressed the time-efficiency of NG-IVIM by investigating several b-value sets in the context of their effect on parameter estimation precision, and we showed that for NG-IVIM increasing the number of b-values beyond 15 did not improve the repeatability of NG-IVIM parameters. Therefore, we could save up to 50% of the scan time without losing precision. So, longer scan time does not always imply better image quality or parameter quantification and optimization of key acquisition parameters is valuable for obtaining time-efficient acquisitions. Time-efficiency is an important criterion for optimizing quantitative imaging, because shorter acquisitions limit the strain on the patient, and on study and clinical resources. Additionally, shorter acquisitions typically contain less patient motion, which is beneficial for precise and accurate parameter estimation.

Single-center clinical validation

In **Chapter 5** we work towards single-center clinical validation of NG-IVIM by incorporating it into the study protocol of the COMPLETE study, where tumor response in HPV-negative oropharyngeal cancer patients will be assessed. Besides NG-IVIM, ex vivo response of tumor biopsies to represent microscopic response [235], and circulating tumor DNA to represent molecular response [143, 145] are assessed as response predictors in the COMPLETE study. By including biomarkers able to measure different response characteristics of the tumor, we can provide a complete, holistic picture of tumor response. The results of this study could in turn give rise to larger studies on NG-IVIM as a response predictor.

In **Chapter 6**, we show that a low pretreatment diffusion coefficient from NG-IVIM indicates a poor response one year after radiotherapy in the subgroup of HPV-negative patients. However, we also show that HPV status has a significant

effect on all NG-IVIM parameters. In fact, our results indicate that the diffusion coefficient from NG-IVIM could serve as a surrogate for HPV status (being a strong prognostic factor itself) rather than an independent predictive factor. Therefore, in future studies, it is vital to correct for HPV status when assessing the predictive power of NG-IVIM. Nevertheless, the response to treatment within HPV-negative patients is still heterogeneous and we show this could be detected by NG-IVIM. Therefore, the results from **Chapter 6** demonstrate the potential added value of NG-IVIM in addition to HPV status for response prediction.

9.2.2 Multi-delay pCASL

Single-center technical development and validation

In **Chapter 3** we took the first step towards using multi-delay pCASL in the head and neck, as only single-delay pCASL was used previously. Multi-delay pCASL enables us to also estimate arterial transit time and therefore correct for any bias in blood flow quantification due to arterial transit time, providing more accurate blood flow measurements. We extensively assessed the nominal values and repeatability, providing reference values for the first application of multi-delay pCASL in head and neck cancer patients.

Now we have shown that acquiring multi-delay pCASL is feasible in the head and neck, a next step to take is to select the most optimal post-labeling delays and labeling durations to ensure time-efficient acquisition in terms of both technical and clinical optimality, and to study the potential clinical value of multi-delay pCASL as a response predictor. Additionally, current studies, including our study in **Chapter 3**, used the blood-tissue partition coefficient (an important parameter for blood flow quantification) also used in brain studies. Future studies should research the blood-tissue partition coefficient specifically for tissues in the head and neck, as this would likely yield a more accurate blood flow quantification. Once these optimal settings have been established, they could be used as a basis of standardization needed for translation to a multi-center research tool (O'Connor domain 2, steps 3 and 4).

In **Chapter 3** we chose to do multi-delay pCASL at 3T, where it was feasible to obtain acceptable image quality in around 5 minutes, which was twice as fast as at 1.5T which means limiting time for patient movement and smaller time slots per patient. However, optimization of post-labeling delays and labeling times for multi-delay pCASL as mentioned in the previous paragraph, or even selecting the optimal post-labeling delay and labeling duration of single-delay pCASL, could potentially also make pCASL of the head and neck feasible at 1.5T within acceptable scan time.

Finally, both NG-IVIM and multi-delay pCASL provide information on perfusion and are theoretically related to each other [112]. Although a single quantitative MRI acquisition could provide essential information on treatment response in under 5 minutes, adding multiple acquisitions providing a variety of biomarkers may improve the accuracy of response prediction. However, this can cause scan protocols to become lengthy and put significant strain on the patient, and the available MRI scanning and analysis capacity. To avoid unnecessary usage of study and clinical resources, it was important to evaluate whether NG-IVIM and multi-delay pCASL provide the same information, in which case acquiring both would have no added value. Therefore, in **Chapter 4** we investigated the correlation between multi-delay pCASL and IVIM in the submandibular glands, parotid glands, and tonsils of healthy volunteers. Since we found no correlation between multi-delay pCASL and IVIM in these healthy head and neck tissue, our results indicate that multi-delay pCASL and IVIM describe different aspects of perfusion. Therefore, investigating both in terms of response prediction could be valuable.

9.2.3 General considerations

The final goal for NG-IVIM and multi-delay pCASL used in **Chapters 2-4** is to obtain measures of response pretreatment or early during treatment, which can guide treatment of individual patients. While the first part of this thesis mainly focused on tumor response, the same techniques could be employed to monitor response of OARs.

A major aspect of assessing response early during treatment is selecting the right time point(s) to measure during treatment. The introduction of the MR-linac, a combination of an MRI scanner with a linear accelerator used for radiotherapy [236], has given us a powerful tool to assess this. With the MR-linac, promising biological MR biomarkers could be acquired each day in the time period where online plan adaptation is done. Since the MR-linac allows for daily imaging, future studies with NG-IVIM and multi-delay pCASL on the MR-linac can give us insight into whether MR parameters actually change during treatment, and if they do change; when they change during treatment. In this way, optimal time-points to acquire biomarkers for plan adaptation can be defined for each treatment schedule.

9.3 Better understanding of treatment-related side effects: osteoradionecrosis

Patients with head and neck cancer who receive radiotherapy typically suffer from one or multiple side effects, for example, a dry mouth (xerostomia) and difficulty swallowing (dysphagia). Osteoradionecrosis (ORN) is a severe late

side effect mainly affecting the mandible after radiotherapy of the head and neck, but it is relatively rare. Therefore, dose-response models for ORN are not as well established as for other, more common side effects.

In **Chapter 7** we proposed the Equivalent Uniform Dose (EUD) as an alternative to conventional DVH parameters that is a suitable variable in dose-response modeling for both conventional and hypofractionated schemes. Conventionally dose-volume histogram (DVH) parameters are used in dose-response models to account for delivered dose. However, conventional DVH parameters are not suitable for describing dose distributions of hypofractionated stereotactic boost schemes. We show that pretreatment teeth extractions, mandibular volume, and the EUD proposed in this work contribute significantly to the risk of ORN. Given the renewed interest in hypofractionation in the field, the EUD could potentially be used as a dose parameter in combined cohorts from multiple centers or when including both historic and modern cohorts with different fractionation schemes. This enables inclusion of larger patient groups needed to establish an accurate dose-response model for osteoradionecrosis.

In **Chapter 8**, we show that there is a significant local effect of teeth extractions and dose. Therefore, a local dose-response model could be a valuable tool in guiding treatment planning once independently validated. Future studies should take into account local radiosensitivity in order to yield a more accurate prediction of ORN. The local measure of dose presented in **Chapter 8** addresses an important shortcoming of conventional DVH parameters. Namely that conventional DVH parameters do not take into account the spatial heterogeneity of the dose. While this might not be an issue for small OARs, larger OARs like the mandible generally have significant heterogeneity in their dose distributions. Moreover, we showed that this heterogeneity locally affects the probability of ORN in the mandible.

In **Chapter 8**, we used generalized estimating equations [227, 228] as a robust method to be able to account for correlations within a patient in a relatively small dataset. Such an approach could also be used for modeling voxel-wise response from imaging-based predictors. In that case, we need to take into account the correlation between voxels, since it is plausible that voxels within one tissue of one patient correlate significantly, similar to how different mandibular regions within one patient correlate significantly.

Although the focus and preference of the scientific community is often on prospective studies, a wealth of data is being acquired from patients that are not part of clinical studies. As we demonstrate in **Chapters 7 and 8**, we can still keep learning from the patients that were already treated. In this case,

however, it is vital that biomarkers and/or patient response and toxicity scoring is done in a systematic and standardized way, to ensure valid conclusions from these studies.

9.4 Future directions

Once we have a validated model available in the clinic for response prediction (O'Connor domain 3), for example, an independently validated model based on the results from **Chapter 5** or **Chapter 8**, the next step is to use the resulting prediction in the clinic. While the simplest way to use a prediction measure is to more accurately inform the patient, a prediction model could be much more valuable if used for the adjustment of treatment, either based on pretreatment measures, or on measures acquired early during treatment. In this way, we may achieve a better response for those tumors resistant to treatment, but also reduce side effects in patients with tumors that respond well to treatment. The personalization approaches mentioned in Figure 9.1 may involve new treatment planning objectives for OARs, or dose (de-)escalation. In the next subsections, several considerations for treatment planning and dose (de-)escalation in personalized radiotherapy are highlighted.

Treatment planning and delivery

A major aspect to take into account is that the feasibility of treatment adjustments depends on the technical (im)possibilities of our treatment techniques. In radiotherapy, we rely on radiation transmitted through other tissues to reach the tumor. These other, healthy tissues unavoidably receive dose, and depending on the function and sensitivity of the tissue this dose will give rise to toxicities. Although in the current clinical workflow many OARs have already been identified and considerable effort is going into lowering the dose in these OARs on a patient-to-patient basis, the radiation has to enter the body somewhere. The model presented in **Chapter 8** could lead to novel, personalized and local dose objectives for the mandible. However, steering away dose from the mandible might result in a higher dose in the salivary glands, oral cavity, and/or swallowing muscles, and radiotherapy treatment planning remains a careful balance of all treatment objectives. So, as multiple new dose objectives emerge, we will find ourselves in a situation where we cannot satisfy them all, as this currently is already the case for some patients. Studies into the feasibility of novel dose objectives therefore could be valuable. Additionally, we could consider employing a different treatment approach to be able to reach the objectives, such as proton irradiation instead of photon irradiation, in which generally less healthy tissue is irradiated due to the fact we make use of the characteristics of protons to stop in the tissue. Another approach could be to use techniques that could widen the therapeutic window

by sensitizing the tumor such as hyperthermia [237], or to take advantage of differences in biological sensitivity between tumor and normal tissues to ultrahigh dose rate irradiation such as in FLASH radiotherapy [238].

Dose (de-)escalation

The main concern of dose escalation is the increased dose to OARs, which could lead to increased toxicity rates. Part of this might be mitigated through treatment planning, by careful balancing of objectives, or (online) adaptive radiotherapy which might allow for smaller PTV margins, as well as selecting patients that are less prone to toxicities for boosting. However, there will be a threshold point in dose escalation at which we cannot avoid overdosage of the organs at risk. Several studies have investigated dose escalation in head and neck tumors which are summarized in a review by Atwell et al. [239]. They indicate that dose escalation in locally advanced head and neck cancer could be feasible, perhaps even up to 10 Gy (EQD2Gy, $\alpha/\beta=10$ Gy), although caution is advised. Instead of boosting the whole tumor, which could lead to a significant increase in integral body dose and dose to OARs, we could also select those regions within the target area with poor prognosis with the help of predictive maps of the tumor based on biological imaging. By selectively boosting, we could dose escalate areas that require a higher dose, but keep the increase in integral body dose to a minimum.

The main concern of treatment de-escalation is under treating the tumor, since eventually, a minimum dose needs to be delivered to still have acceptable tumor control and overall or progression free survival. Many studies on dose de-escalation based on HPV status are currently ongoing [240]. One strategy for de-escalation in radiotherapy could consist of changing the total dose and/or number of fractions. Another strategy could be to decrease the irradiated volume, for example by omitting irradiation of lymph nodes [241].

Other options for treatment (de-)escalation

Since treatment of head and neck cancer is multidisciplinary, approaches other than changing radiotherapy can be taken for treatment (de-)escalation. Several studies investigated treatment de-escalation in HPV-positive patients by changing the concurrent chemotherapy agent from cisplatin to cetuximab [242, 243], however, they demonstrated significantly worse tumor control, overall survival, and progression free survival with no added benefit in terms of reduced toxicity. Another method for treatment escalation could be to add immunotherapy to concurrent chemoradiotherapy [244].

9.5 Final remarks

The key to personalization of radiotherapy for head and neck cancer lies in the ability to develop accurate models to predict tumor response and toxicity. One aspect of this is establishing markers capable of predicting tumor response pretreatment and/or early during treatment. In this thesis, we therefore evaluated and optimized two quantitative MRI techniques, NG-IVIM and multi-delay pCASL, and made valuable steps towards applying them in larger groups of patients and/or a multi-center setting for clinical validation in terms of response prediction. Additionally, we presented a study protocol to assess the predictive value of NG-IVIM in HPV-negative patients. In terms of improved prediction of ORN risk, we opened the door for multi-center dose-response modeling of ORN of the mandible by presenting the EUD as a dose parameter suitable for different fractionation schemes. Besides that, we provided a local dose-response model showing ORN of the mandible depends locally on dose, teeth extractions, and smoking. Finally, after thorough validation, the MRI techniques and dose-response models presented in this thesis could be used in treatment planning and decision-making processes, which could lead to improved patient outcomes through personalization of radiotherapy.

S

Summary
Samenvatting

Summary

Tumor response to treatment is heterogeneous among patients with head and neck cancer. At the same time, head and neck radiation therapy is complex due to the many organs at risk (OARs) involved and irradiation can lead to significant toxicity in treated patients. So, a considerable group of patients does not respond well to treatment, while at the same time suffering from side effects.

Personalization of treatment could improve response to treatment while minimizing side effects, for example through personalization approaches such as (1) more dose to tumors that need more dose, (2) less dose to tumors that need less dose, (3) higher biological dose, (4) more dose to patients that can tolerate more dose, (5) selective sparing of complication prone regions. To employ any of these personalization approaches, methods for response and toxicity prediction are needed. Therefore, the primary objective of this thesis was to develop imaging-based markers that provide insight into prediction of tumor response and treatment-related toxicity. The aim is to use these predictors for optimization of treatment planning and decision-making, ultimately leading to improved patient outcomes.

Tumor response is multi-factorial, and predictors describing different aspects of tumor response are likely needed to obtain a complete picture of tumor response (**Chapter 5**). Therefore, the first part of this thesis focuses on prediction of tumor response using MRI techniques providing information on perfusion and diffusion properties of tissue: non-Gaussian intravoxel incoherent motion (NG-IVIM) imaging and multi-delay pseudo-continuous arterial spin labeling (pCASL). On the other hand, OAR response to treatment could also differ per patient, with some patients tolerating more dose than other patients. Therefore, a key aspect of treatment personalization is establishing personalized dose objectives for OARs. The second part of this thesis focuses on improving our understanding of treatment-related toxicity, specifically for osteoradionecrosis (ORN) of the mandible.

Part 1: MRI for tumor response prediction

The first part of this thesis focuses on using MRI for tumor response prediction. In this thesis, NG-IVIM DWI and multi-delay pCASL were investigated as promising candidates for response prediction biomarkers.

NG-IVIM DWI is a versatile technique that can yield parameters related to diffusion, cellularity, and perfusion of the tissue on the micro-environment level. These are typically characteristics that are disrupted in tumors and that could

change during treatment, which makes NG-IVIM a promising technique for response prediction. In **Chapter 2** we optimized diffusion-weighting b-values, a key acquisition parameter of DWI, in b-value set sizes ranging between 5 and 30 b-values by optimizing the Cramér-Rao Lower Bound for the NG-IVIM model. We assessed the effect of the number of b-values, and the influence of the optimization on parameter estimation precision. In 16 healthy subjects, we showed that b-value optimization is vital for time-efficient acquisition of NG-IVIM and presented a b-value set of 15 b-values with comparable precision to a b-value set of 30 b-values, with a 50% reduction in scan time. In terms of post-processing, we investigated three registration types: none versus inter-volume versus intra- and inter-volume registration, which all yielded similar parameter estimation precision. Additionally, we investigated the effect of manual swallowing artifact rejection and showed rejection of images containing swallowing artifacts removed the bias in parameter estimation if present. This optimized acquisition was implemented in the clinical study described in **Chapter 5**.

ASL is a technique that has only recently gained interest in the head and neck region. It is capable of measuring blood flow to tissues without needing the administration of external contrast agent. Because perfusion is typically disrupted in tumors, ASL is also an interesting candidate for response prediction. Previous studies have used single-delay pCASL in the head and neck, however, using multi-delay pCASL has the advantage of providing accurate blood flow measurements in multiple tissues simultaneously. In **Chapter 3** we demonstrated the feasibility of using multi-delay pCASL in the head and neck. Since little is known about the performance and characteristics of ASL in the head and neck, we also explored nominal values and repeatability of the blood flow in the submandibular glands, parotid glands, and tonsils of 20 healthy volunteers. The nominal blood flow (\pm SE) was 55.7 ± 3.1 ml/100g /min for the parotid glands, 41.2 ± 2.8 ml/100 g/min for the submandibular glands, and 32.3 ± 2.2 ml/100 g/min for the tonsils. The best repeatability was found in the parotid glands (within-subject coefficient of variation (wCV) = 13.3%–16.1%), followed by the submandibular glands and tonsils (wCV=20.0–24.6%). These values can serve as a reference for future studies investigating the clinical value of multi-delay pCASL in the head and neck.

Since both IVIM (which estimates the same perfusion-related parameters as NG-IVIM) and multi-delay pCASL provide perfusion-related measurements of the underlying tissue, it would be unnecessary to acquire both if they provide the same perfusion-related information. Therefore, in **Chapter 4** we studied the correlation between blood flow and arterial transit time obtained from multi-delay pCASL, and perfusion fraction and the pseudo-diffusion coefficient

from IVIM in the same cohort as **Chapter 3**. We found no strong or significant correlations in the submandibular glands, parotid glands, and tonsils, indicating that IVIM and multi-delay pCASL describe different aspects of perfusion and are both worth investigating in the context of response prediction.

Chapter 5 describes the COMPLETE study, the prospective clinical study we started to investigate novel prediction markers in human papillomavirus (HPV)-negative oropharyngeal squamous cell carcinoma (OPSCC) patients. Patients included in the study are scanned with the NG-IVIM protocol, described in **Chapter 2**, before, during, and after treatment. The COMPLETE trial aims to provide a complete, holistic view of the tumor landscape by accounting for different tumor layers; NG-IVIM to measure the macroscopic tumor response, ex vivo radiation of tissue culture to measure the microscopic tumor response, and circulating tumor DNA to measure the molecular tumor response. Thereby gaining insight into the value of novel potential biomarkers acquired before or during treatment for the prediction of response and locoregional control two years after radiotherapy in HPV-negative OPSCC.

Chapter 6 describes the first use of the optimized NG-IVIM acquisition presented in **Chapter 2** in a group of 36 OPSCC patients. We showed that in a subgroup analysis of HPV-negative patients a lower diffusion coefficient was significantly related to poorer response to treatment. This was an unexpected finding, since previous studies report that a lower diffusion coefficient, and HPV-positive status, are associated with favorable response. In this study, we also found HPV status has a significant effect on the apparent diffusion coefficient ADC , and all NG-IVIM parameter values. Specifically, HPV-positive tumors had a higher pseudo-diffusion coefficient D^* and kurtosis K , and a lower diffusion coefficient D and perfusion fraction f , and ADC than HPV-negative tumors. Furthermore, we constructed clusters based on NG-IVIM parameter values representing different intra-tumor phenotypes, and found the cluster with a high diffusion coefficient D was significantly less present in HPV-positive tumors compared to HPV-negative tumors. So, our results indicate that the effect of HPV status on diffusion coefficient D can obscure the true predictive value of D , and render D merely surrogates of HPV status. Therefore, it is important to correct for HPV status, or analyze HPV-positive and HPV-negative groups separately when assessing the predictive value of NG-IVIM in OPSCC. We found no correlation between response and ADC , indicating the potential added value of the more elaborate NG-IVIM model compared to conventional DWI. **Chapter 6** concludes the first part of the thesis.

Part 2: Response prediction of organs at risk

For personalization patients can be stratified either based on tumor response prediction, expected OAR tolerance, or both. The second part of this thesis is on response prediction in OARs. Specifically, it focuses on the prediction of ORN in the mandible, as severe radiotherapy-related treatment toxicity in the head and neck. Because ORN of the mandible is a relatively rare toxicity, establishing a dose-response model requires a large group of patients, which may have different fractionation schemes, especially in light of renewed interest in hypofractionation. In **Chapter 7** we showed that conventional dose-volume histogram (DVH) parameters (for example V70Gy, the volume receiving 70 Gy), do not perform well for hypofractionated treatment. To avoid needing to build a model for each separate fractionation scheme, we proposed the equivalent uniform dose (EUD) as a suitable alternative for conventional DVH parameters to arrive at a single model that could be used across fractionation schemes. The dose-response model presented in **Chapter 7** was built using data from 334 patients of which 32 developed ORN. We showed that the EUD with $\alpha=8$ has the best discrimination between ORN and non-ORN patients (AUC=0.71). In multivariate analyses, pre-radiotherapy extractions (subhazard ratio (SHR) = 2.34; $p = 0.012$), mandibular volume (SHR = 1.04; $p = 0.003$), and the EUD ($\alpha=8$) (SHR = 1.14; $p < 0.001$) were significantly associated with ORN in a single model that can be used for both conventionally fractionated schemes and hypofractionated stereotactic boost schemes.

Not all regions of the mandible may be equally at risk of ORN due to heterogeneity of the dose distribution and local trauma to the mandible in the form of teeth extractions. Therefore, in **Chapter 8** we built a local dose-response model for ORN of the mandible. For this purpose, we divided the mandible into 16 volumes of interest (VOIs) based on the dental elements. Of the 219 patients included, 22 developed ORN in 89 element VOIs. We applied generalized estimating equations to build a model while also taking into account the correlation between element VOIs of the same patient. Finally, mean dose to the element VOI (odds ratio (OR)=1.05 per Gy, 95% confidence interval (CI): (1.04,1.07)), pre-radiotherapy extractions of an element ipsilateral to element of interest (including extraction of the element of interest itself) (OR=2.81, 95% CI: (1.12,7.05)), and smoking at start of radiotherapy (OR=3.37, 95% CI: (1.29,8.78)) were significantly associated with an increased probability of ORN in the VOI. Our results indicate that the probability of developing ORN varies within the mandible and strongly depends on the location of teeth extractions, local dose, and smoking.

The thesis ends with the general discussion presented in **Chapter 9**, where the findings of this thesis are discussed in a broader context and in light of future

application to personalization of head and neck radiotherapy. In conclusion, the key to personalization of radiotherapy for head and neck cancer lies in the ability to develop accurate models to predict tumor response and toxicity. The work presented in this thesis contributes to improving the accuracy of response prediction by (i) providing improved MRI techniques to measure response before and early during treatment and (ii) by providing dose-response models for ORN of the mandible incorporating key spatial information and the EUD as a generalizable dose parameter across different fractionation schemes. Both can now be validated in larger groups of patients, after which they could contribute to personalized treatment planning and decision making processes.

Samenvatting

Tumorrespons op behandeling is heterogeen in patiënten met hoofd-hals kanker. Tegelijkertijd is radiotherapie van het hoofd-hals gebied complex, omdat er veel gezonde organen betrokken zijn en bestraling van deze risico-organen kan zorgen voor significante bijwerkingen. Een aanzienlijke groep patiënten reageert niet goed op de behandeling, terwijl ze ook lijden aan verschillende bijwerkingen.

Personalisatie van de behandeling kan de respons op de behandeling verbeteren en tegelijkertijd het risico op bijwerkingen minimaliseren, bijvoorbeeld door strategieën als (1) meer dosis voor tumoren die meer dosis nodig hebben, (2) minder dosis voor tumoren die minder dosis nodig hebben, (3) hogere biologische dosis, (4) meer dosis voor patiënten die meer dosis kunnen tolereren, en (5) selectief sparen van regio's die een groot risico op complicaties hebben. Om deze strategieën te kunnen gebruiken, zijn er eerst methodes nodig om tumorrespons en bijwerkingen te kunnen voorspellen. Daarom was de primaire doelstelling van deze thesis om op beeldvorming gebaseerde markers te ontwikkelen die inzicht geven in de respons van de tumor en aan de behandeling gerelateerde bijwerkingen. Het streven is om deze voorspellers te gebruiken voor het optimaliseren van het radiotherapie behandelplan en besluitvormende processen, wat uiteindelijk leidt tot verbeterde uitkomsten voor de patiënt.

Tumorrespons hangt af van verschillende factoren, en voorspellers die verschillende aspecten van tumorrespons kunnen beschrijven zijn hoogstwaarschijnlijk nodig om een compleet beeld van tumorrespons te verkrijgen (**Hoofdstuk 5**). Daarom richt het eerste deel van deze thesis zich op voorspelling van tumorrespons met MRI technieken die informatie geven over perfusie en diffusie eigenschappen van weefsel: non-Gaussian intravoxel incoherent motion (NG-IVIM) diffusie-gewogen beeldvorming (DWI) en multi-delay pseudo-continue arterial spin labeling (pCASL). Aan de andere kant kan respons van risico-organen ook verschillen per patiënt, waarbij sommige patiënten meer dosis kunnen tolereren dan andere patiënten. Een belangrijk aspect van personalisatie van de behandeling is daarom het vaststellen van gepersonaliseerde dosis (streef)limieten voor risico-organen. Het tweede deel van deze thesis richt zich daarom op het beter begrijpen van behandeling gerelateerde bijwerkingen, specifiek voor osteoradionecrose (ORN) van de onderkaak.

Deel 1: MRI voor tumorresponsvoorspelling

Het eerste deel van deze thesis richt zich op het gebruik van MRI voor de voorspelling van tumorrespons. In deze thesis zijn NG-IVIM en multi-delay pCASL onderzocht als veelbelovende kandidaten voor respons voorspellende biomarkers.

NG-IVIM DWI is een veelzijdige techniek wiens parameters informatie geven over diffusie, cellulariteit en perfusie van het weefsel op het niveau van de tumor micro-omgeving. Dit zijn typisch eigenschappen die verstoord zijn in tumoren, en die tijdens de behandeling kunnen veranderen, wat NG-IVIM een veelbelovende techniek voor responsvoorspelling maakt. In **Hoofdstuk 2** hebben we de diffusie-gewogen b-waardes, een belangrijke acquisitie parameter van DWI, geoptimaliseerd op basis van de Cramér-Rao lower bound voor het NG-IVIM model, in set groottes van 5 tot 30 b-waardes. We hebben het effect van het aantal b-waardes, en de invloed van de optimalisatie op de precisie van parameter schatting geëvalueerd. In 16 gezonde vrijwilligers hebben we laten zien dat optimalisatie essentieel is voor tijdsefficiënte acquisitie van NG-IVIM. Daarnaast hebben we een b-waardes set van 15 b-waardes gepresenteerd die vergelijkbare precisie vertoont als een set van 30 b-waardes, maar in de helft van de scantijd. Daarnaast hebben we het effect van drie verschillende registratie types bekeken: geen versus alleen inter-volume registratie versus intra- en inter-volume registratie. Alle drie deze registratie types lieten vergelijkbare parameter precisie zien. Ook hebben we het effect van handmatige verwijderen van beelden met slikartefacten onderzocht, en lieten zien dat dit de bias in parameter schatting wegneemt. De gepresenteerde acquisitie werd geïmplementeerd in het klinische studie protocol van **Hoofdstuk 5**.

ASL is een techniek die pas recentelijk aandacht heeft gekregen in het hoofd-hals gebied. ASL kan de bloedstroom naar weefsels meten zonder dat hiervoor contrastvloeistof toegediend hoeft te worden. Omdat perfusie vaak verstoord is in tumoren, is ook ASL een interessante kandidaat voor respons voorspelling. Eerdere studies in het hoofd-hals gebied maakten gebruik van single-delay pCASL, maar het gebruik van multi-delay pCASL heeft het voordeel dat het accurate metingen van bloedstroom kan doen in verschillende weefsels tegelijk. In **Hoofdstuk 3** hebben we de haalbaarheid van het gebruik van multi-delay pCASL in het hoofd-hals gebied gedemonstreerd. Omdat er weinig bekend is over de prestaties en karakteristieken van ASL in het hoofd-hals gebied, hebben we ook nominale waardes en de herhaalbaarheid van bloedstroom in de onderkaakspeekselklieren, oorspeekselklieren en tonsillen onderzocht in 20 gezonde vrijwilligers. De nominale bloedstroom (\pm standaard fout) was 55.7 ± 3.1 ml/100 g/min voor de oorspeekselklieren, 41.2 ± 2.8 ml/100

g/min voor de onderkaakspeekselklieren, en 32.3 ± 2.2 ml/100 g/min voor de tonsillen. De beste herhaalbaarheid is gevonden in de oorspeekselklieren (binnen subject coëfficiënt van variatie (wCV) = 13.3%-16.1%), gevolgd door de onderkaakspeekselklieren en tonsillen (wCV=20.0-24.6%). Deze waarden kunnen als referentie dienen voor toekomstige studies die de klinische waarde van multi-delay pCASL in het hoofd-hals gebied onderzoeken.

Aangezien zowel IVIM (welke dezelfde perfusie-gerelateerde parameters schat als NG-IVIM) als multi-delay pCASL perfusie-gerelateerde informatie geven over het onderliggende weefsel, zou het niet nodig zijn om beiden te verwerven als ze dezelfde informatie geven. Daarom hebben we in **Hoofdstuk 4** de correlatie tussen bloedstroom en arteriële doorvoertijd van het bloed van multi-delay pCASL, en de perfusie fractie en pseudo-diffusie coëfficiënt van IVIM onderzocht in dezelfde gezonde vrijwilligers als in **Hoofdstuk 3**. We hebben geen sterke of significante correlaties gevonden in de onderkaakspeekselklieren, oorspeekselklieren of de tonsillen, wat suggereert dat IVIM en multi-delay pCASL verschillende aspecten van perfusie beschrijven, en dat beiden het waard zijn om in de context van responsvoorspelling te onderzoeken.

Hoofdstuk 5 beschrijft de COMPLETE studie, de prospectieve klinische studie die we gestart zijn om nieuwe voorspellende markers te onderzoeken in patiënten met human papillomavirus (HPV)-negatief oropharynxcarcinoom. Patiënten die geïncludeerd zijn in de studie worden gescand met het NG-IVIM protocol beschreven in **Hoofdstuk 2**, voor, tijdens en na behandeling. Naast MRI, ambieert de COMPLETE studie om een compleet, holistisch beeld van het tumor landschap te verkrijgen door verschillende tumorlagen mee te nemen: NG-IVIM voor het meten van macroscopische tumorrespons, ex vivo bestraling van weefselcultuur om de microscopische respons te meten, en circulerend tumor DNA om de moleculaire respons te meten. Op deze manier kunnen we inzicht krijgen in de waarde van nieuwe potentiële biomarkers voor de voorspelling van respons en locoregionale controle twee jaar na bestraling van HPV-negatief oropharynxcarcinoom.

Hoofdstuk 6 beschrijft de eerste toepassing van de geoptimaliseerde NG-IVIM acquisitie uit **hoofdstuk 2** in een groep van 36 patiënten met oropharynxcarcinoom. We lieten zien dat in een subgroep van HPV-negatieve patiënten, een lagere diffusie coëfficiënt significant gerelateerd is aan een slechtere respons. Dit was een onverwachte bevinding, aangezien eerdere studies zowel een lagere diffusie coëfficiënt als HPV-positiviteit associëren met een gunstige respons. In deze studie hebben we ook gevonden dat HPV status een significant effect heeft op de apparent diffusion coefficient ADC en alle

NG-IVIM parameter waardes. HPV-positieve tumoren hadden een hogere pseudo-diffusie coëfficiënt D^* en kurtosis K , en een lagere diffusie coëfficiënt D , perfusie fractie f en ADC dan HPV-negatieve tumoren. Verder hebben we clusters op basis van NG-IVIM parameter waardes geconstrueerd. Deze clusters vertegenwoordigen verschillende intra-tumor fenotypes. We zagen dat het cluster met een hoge D significant minder aanwezig was bij HPV-positieve tumoren in vergelijking met HPV-negatieve tumoren. Onze resultaten suggereren dus dat het effect van HPV-status op de D de echte voorspellende waarde van D kan verdoezelen, waardoor de D louter een surrogaat van HPV-status wordt. Daarom is het belangrijk om voor HPV-status te corrigeren als de voorspellende waarde van NG-IVIM in oropharynxcarcinoom onderzocht wordt. We vonden geen correlatie tussen respons en ADC , wat een indicatie geeft van de potentieel toegevoegd waarde van NG-IVIM ten opzichte van conventionele DWI. **Hoofdstuk 6** concludeert deel één van deze thesis.

Deel 2: Responsvoorspelling van risico-organen

Voor personalisatie van de behandeling kunnen patiënten onderverdeeld worden op basis van tumorresponsvoorspelling, of de verwachte tolerantie van risico-organen. Het tweede deel van deze thesis gaat over response voorspelling in risico-organen. Specifiek richt het zich op de voorspelling van ORN in de onderkaak als ernstige bijwerking gerelateerd aan radiotherapie van het hoofd-hals gebied. Omdat ORN van de onderkaak relatief zeldzaam is, zijn grote groepen patiënten nodig om een dosis-respons model te bewerkstelligen. Grote groepen patiënten kunnen verschillende fractioneringsschema's ontvangen, zeker in het licht van de hernieuwde interesse in hypofractionering. In **Hoofdstuk 7** laten we zien dat conventionele dosis-volume histogram (DVH) parameters (zoals de $V70$ Gy, het volume van de onderkaak dat 70 Gy ontvangt), niet goed presteren voor gehypofractioneerde behandelingen. Om te voorkomen dat er losse modellen nodig zijn voor elk fractioneringsschema, stellen wij de equivalente uniforme dosis (EUD) voor als een passend alternatief voor conventionele DVH-parameters om uiteindelijk uit te komen op één model dat voor verschillende fractioneringsschema's gebruikt kan worden. Het dosis-responsmodel uit **Hoofdstuk 7** is gebaseerd op data van 334 patiënten van wie er 32 ORN ontwikkelden. We laten zien dat de EUD met $\alpha=8$ de beste discriminatie oplevert tussen ORN en niet-ORN patiënten (AUC = 0.71). In multivariabele analyse waren pre-radiotherapie tand- en kiesextracties (subhazard ratio (SHR) = 2.34; $p = 0.012$), het volume van de onderkaak (SHR = 1.04, $p=0.003$), en de EUD ($\alpha=8$; SHR=1.14; $p<0.001$) significant geassocieerd met ORN in een enkel model dat gebruikt kan worden voor zowel conventioneel gefractioneerde schema's als gehypofractioneerde bestralingschema's.

Wellicht lopen niet alle regio's van de onderkaak evenveel risico op het ontwikkelen van ORN door heterogeniteit van dosis distributie en lokaal trauma van de onderkaak door tand- en kiesextracties. Daarom hebben we in **Hoofdstuk 8** een lokaal dosis-responsmodel gebouwd voor ORN van de onderkaak. Om dit te bewerkstelligen hebben we de onderkaak opgedeeld in 16 volumes van interesse (VOIs) gebaseerd op de dentale elementen. Van de 219 patiënten die geïnccludeerd zijn, ontwikkelden 22 patiënten ORN in 89 element VOIs. We hebben generalized estimating equations toegepast om een model te bouwen waarin de correlatie tussen de element VOIs van één patiënt meegenomen wordt. Uiteindelijk bleken de gemiddelde dosis op een element VOI (odds ratio (OR)=1.05 per Gy, 95% confidence interval (CI): (1.04,1.07)), pre-radiotherapie extracties van een element ipsilateraal van het element van interesse (inclusief het element van interesse zelf, OR=2.81, 95% CI: (1.12,7.05)), en roken bij de start van radiotherapie (OR=3.37, 95% CI: (1.29,8.78)) significant geassocieerd met een toegenomen kans op het ontwikkelen van ORN in een VOI. Onze resultaten suggereren dus dat de kans op het ontwikkelen van ORN varieert binnen de onderkaak en sterk afhangt van de locatie van tand- en kiesextracties, lokale dosis en roken.

Deze thesis eindigt met een algemene discussie in **Hoofdstuk 9**, waar de bevindingen in deze thesis bediscussieerd worden in een bredere context en in het licht van de toekomstige toepassing in personalisatie van hoofd-hals radiotherapie. Concluderend, de sleutel tot personalisatie van radiotherapie voor hoofd-hals kanker ligt in het vermogen om accurate modellen te ontwikkelen om tumorrespons en bijwerkingen te voorspellen. Het werk wat gepresenteerd is in deze thesis draagt bij aan het verbeteren van de accuraatheid van respons voorspelling door (i) verbeterde MRI technieken om respons voor, en vroeg tijdens de behandeling te meten en (ii) door het verstrekken van dosis-responsmodellen voor ORN in de onderkaak die belangrijke spatiële informatie en de EUD als generaliseerbare dosis variabele voor verschillende fractioneringsschema's bevatten. Beiden kunnen nu gevalideerd worden in grotere groepen patiënten, waarna ze kunnen bijdragen aan gepersonaliseerde radiotherapie behandelplannen en besluitvormende processen.

R

References

References

- [1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71:209–49
- [2] Nederlandse Kankerregistratie (NKR). NKR cijfers. <https://iknl.nl/nkr-cijfers>
- [3] Blot WJ, McLaughlin JK, Winn DM, Austin DF, Greenberg RS, Preston–Martin S, et al. Smoking and drinking in relation to oral and pharyngeal cancer. *Cancer Res.* 1988;48:3282–7
- [4] Mehanna H, Beech T, Nicholson T, El–Hariry I, McConkey C, Paleri V, et al. Prevalence of human papillomavirus in oropharyngeal and nonoropharyngeal head and neck cancer–systematic review and meta–analysis of trends by time and region. *Head Neck.* 2013;35:747–55
- [5] Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, et al. Evidence for a Causal Association Between Human Papillomavirus and a Subset of Head and Neck Cancers. *J Natl Cancer Inst.* 2000;92:709–20
- [6] Gillison ML, D’Souza G, Westra W, Sugar E, Xiao W, Begum S, et al. Distinct risk factor profiles for human papillomavirus type 16–positive and human papillomavirus type 16–negative head and neck cancers. *J Natl Cancer Inst.* 2008;100:407–20
- [7] Fakhry C, Westra WH, Li S, Cmelak A, Ridge JA, Pinto H, et al. Improved survival of patients with human papillomavirus–positive head and neck squamous cell carcinoma in a prospective clinical trial. *J Natl Cancer Inst.* 2008;100:261–9
- [8] Liu C, Mann D, Sinha UK, Kokot NC. The molecular mechanisms of increased radiosensitivity of HPV–positive oropharyngeal squamous cell carcinoma (OPSCC): an extensive review. *J Otolaryngol Head Neck Surg.* 2018;47:59
- [9] Wang MB, Liu IY, Gornbein JA, Nguyen CT. HPV–Positive Oropharyngeal Carcinoma: A Systematic Review of Treatment and Prognosis. *Otolaryngol Head Neck Surg.* 2015;153:758–69
- [10] Liu X, Liu P, Chernock RD, Lang Kuhs KA, Lewis JS, Jr, Li H, et al. Impact of human papillomavirus on the tumor microenvironment in oropharyngeal squamous cell carcinoma. *Int J Cancer.* 2022;150:521–31
- [11] de Perrot T, Lenoir V, Domingo Ayllon M, Dulguerov N, Pusztaszeri M, Becker M. Apparent Diffusion Coefficient Histograms of Human Papillomavirus–Positive and Human Papillomavirus–Negative Head and Neck Squamous Cell Carcinoma: Assessment of Tumor Heterogeneity and Comparison with Histopathology. *AJNR Am J Neuroradiol.* 2017;38:2153–60
- [12] Uzun S, Korkmaz Y, Wuerdemann N, Arolt C, Puladi B, Siefer OG, et al. Comprehensive Analysis of VEGFR2 Expression in HPV–Positive and –Negative OPSCC Reveals Differing VEGFR2 Expression Patterns. *Cancers.* 2021;13:5221
- [13] Amin MB, Greene FL, Edge SB, Compton CC, Gershengwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population–based to a more “personalized” approach to cancer staging. *CA Cancer J Clin.* 2017;67:93–9
- [14] Gupta T, Ghosh–Laskar S, Agarwal JP. Resource–sparing curative–intent hypofractionated–accelerated radiotherapy in head and neck cancer: More relevant than ever before in the COVID era. *Oral Oncol.* 2020;111:105045

- [15] Thomson DJ, Palma D, Guckenberger M, Balermpas P, Beitler JJ, Blanchard P, et al. Practice Recommendations for Risk-Adapted Head and Neck Cancer Radiation Therapy During the COVID-19 Pandemic: An ASTRO-ESTRO Consensus Statement. *Int J Radiat Oncol Biol Phys.* 2020;107:618-27
- [16] Huang SH, O'Sullivan B, Su J, Ringash J, Bratman SV, Kim J, et al. Hypofractionated radiotherapy alone with 2.4 Gy per fraction for head and neck cancer during the COVID-19 pandemic: The Princess Margaret experience and proposal. *Cancer.* 2020;126:3426-37
- [17] Dirix P, Nuyts S, Van den Bogaert W. Radiation-induced xerostomia in patients with head and neck cancer: a literature review. *Cancer.* 2006;107:2525-34
- [18] Garcia-Peris P, Paron L, Velasco C, de la Cuerda C, Cambolor M, Breton I, et al. Long-term prevalence of oropharyngeal dysphagia in head and neck cancer patients: Impact on quality of life. *Clin Nutr.* 2007;26:710-7
- [19] Gunn L, Gilbert J, Nenclares P, Soliman H, Newbold K, Bhide S, et al. Taste dysfunction following radiotherapy to the head and neck: A systematic review. *Radiother Oncol.* 2021;157:130-40
- [20] Sapir E, Tao Y, Feng F, Samuels S, El Naqa I, Murdoch-Kinch CA, et al. Predictors of Dysgeusia in Patients With Oropharyngeal Cancer Treated With Chemotherapy and Intensity Modulated Radiation Therapy. *Int J Radiat Oncol Biol Phys.* 2016;96:354-61
- [21] Vainshtein JM, Samuels S, Tao Y, Lyden T, Haxer M, Spector M, et al. Impact of xerostomia on dysphagia after chemotherapy-intensity-modulated radiotherapy for oropharyngeal cancer: Prospective longitudinal study. *Head Neck.* 2016;38 Suppl 1:E1605-12
- [22] Baudalet M, Van den Steen L, Tomassen P, Bonte K, Deron P, Huvenne W, et al. Very late xerostomia, dysphagia, and neck fibrosis after head and neck radiotherapy. *Head Neck.* 2019;41:3594-603
- [23] Frankart AJ, Frankart MJ, Cervenka B, Tang AL, Krishnan DG, Takiar V. Osteoradionecrosis: Exposing the Evidence Not the Bone. *Int J Radiat Oncol Biol Phys.* 2021;109:1206-18
- [24] Jellema AP, Slotman BJ, Doornaert P, Leemans CR, Langendijk JA. Impact of radiation-induced xerostomia on quality of life after primary radiotherapy among patients with head and neck cancer. *Int J Radiat Oncol Biol Phys.* 2007;69:751-60
- [25] Dirix P, Nuyts S. Evidence-based organ-sparing radiotherapy in head and neck cancer. *Lancet Oncology.* 2010;11:85-91
- [26] Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* 2022;12:31-46
- [27] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144:646-74
- [28] Hanahan D, Weinberg RA. The Hallmarks of Cancer. *Cell.* 2000;100:57-70
- [29] Bernstein JM, Homer JJ, West CM. Dynamic contrast-enhanced magnetic resonance imaging biomarkers in head and neck cancer: potential to guide treatment? A systematic review. *Oral Oncol.* 2014;50:963-70
- [30] Noij DP, de Jong MC, Mulders LGM, Marcus JT, de Bree R, Lavini C, et al. Contrast-enhanced perfusion magnetic resonance imaging for head and neck squamous cell carcinoma: a systematic review. *Oral Oncol.* 2015;51:124-38
- [31] Yu L, Li C, Luo X, Zhou J, Zhang C, Zhang Y, et al. Differentiation of Malignant and Benign Head and Neck Tumors with Amide Proton Transfer-Weighted MR Imaging. *Mol Imaging Biol.* 2019;21:348-55

References

- [32] Yuan J, Chen S, King AD, Zhou J, Bhatia KS, Zhang Q, et al. 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 Biomed.* 2014;27:1239-47
- [33] Wang J, Weygand J, Hwang KP, Mohamed ASR, Ding Y, Fuller CD, et al. Magnetic Resonance Imaging of Glucose Uptake and Metabolism in Patients with Head and Neck Cancer. *Sci Rep.* 2016;6:30618
- [34] Fujima N, Andreu-Arasa VC, Barest GD, Srinivasan A, Sakai O. Magnetic Resonance Spectroscopy of the Head and Neck: Principles, Applications, and Challenges. *Neuroimaging Clin N Am.* 2020;30:283-93
- [35] Stejskal EO, Tanner JE. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time Dependent Field Gradient. *J Chem Phys.* 1965;42:288-92
- [36] Vandecaveye V, Dirix P, De Keyzer F, de Beeck KO, Vander Poorten V, Roebben I, et al. Predictive value of diffusion-weighted magnetic resonance imaging during chemoradiotherapy for head and neck squamous cell carcinoma. *Eur Radiol.* 2010;20:1703-14
- [37] Kim S, Loevner L, Quon H, Sherman E, Weinstein G, Kilger A, et al. Diffusion-weighted magnetic resonance imaging for predicting and detecting early response to chemoradiation therapy of squamous cell carcinomas of the head and neck. *Clin Cancer Res.* 2009;15:986-94
- [38] King AD, Mo FKF, Yu KH, Yeung DKW, Zhou H, Bhatia KS, et al. Squamous cell carcinoma of the head and neck: diffusion-weighted MR imaging for prediction and monitoring of treatment response. *Eur Radiol.* 2010;20:2213-20
- [39] Lima M, Le Bihan D. Clinical Intravoxel Incoherent Motion and Diffusion MR Imaging: Past, Present, Future. *Radiology.* 2016;278:13-32
- [40] Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of Diffusion and Perfusion in Intravoxel Incoherent Motion MR Imaging. *Radiology.* 1988;168:497-505
- [41] Le Bihan D, Breton E, Lallemand D, Grenier P, Cabanis E, Laval-Jeantet M. MR Imaging of Intravoxel Incoherent Motions: Application to Diffusion and Perfusion in Neurologic Disorders. *Radiology.* 1986;161:401-7
- [42] Lu Y, Jansen JFA, Mazaheri Y, Stambuk HE, Koutcher JA, Shukla-Dave A. Extension of the intravoxel incoherent motion model to non-gaussian diffusion in head and neck cancer. *J Magn Reson Imaging.* 2012;36:1088-96
- [43] Schakel T, Hoogduin JM, Terhaard CHJ, Philippens MEP. Diffusion weighted MRI in head-and-neck cancer: geometrical accuracy. *Radiother Oncol.* 2013;109:394-7
- [44] Koyasu S, Lima M, Umeoka S, Morisawa N, Porter DA, Ito J, et al. The clinical utility of reduced-distortion readout-segmented echo-planar imaging in the head and neck region: initial experience. *Eur Radiol.* 2014;24:3088-96
- [45] Walter SS, Liu W, Stemmer A, Martirosian P, Nikolaou K, Notohamprodjjo M, et al. Combination of integrated dynamic shimming and readout-segmented echo planar imaging for diffusion weighted MRI of the head and neck region at 3 Tesla. *Magn Reson Imaging.* 2017;42:32-6
- [46] Kim TH, Baek MY, Park JE, Ryu YJ, Cheon JE, Kim IO, et al. Comparison of DWI Methods in the Pediatric Brain: PROPELLER Turbo Spin-Echo Imaging Versus Readout-Segmented Echo-Planar Imaging Versus Single-Shot Echo-Planar Imaging. *AJR Am J Roentgenol.* 2018;210:1352-8

- [47] Andersson JLR, Skare S, Ashburner J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage*. 2003;20:870-88
- [48] Winter RM, Schmidt H, Leibfarth S, Zwirner K, Welz S, Schwenzer NF, et al. Distortion correction of diffusion-weighted magnetic resonance imaging of the head and neck in radiotherapy position. *Acta Oncol*. 2017;56:1659-63
- [49] Le Bihan D, Poupon C, Amadon A, Lethimonnier F. Artifacts and pitfalls in diffusion MRI. *J Magn Reson Imaging*. 2006;24:478-88
- [50] Alsop DC, Detre JA, Golay X, Gunther M, Hendrikse J, Hernandez-Garcia L, et al. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magn Reson Med*. 2015;73:102-16
- [51] Fujima N, Kudo K, Yoshida D, Homma A, Sakashita T, Tsukahara A, et al. Arterial spin labeling to determine tumor viability in head and neck cancer before and after treatment. *J Magn Reson Imaging*. 2014;40:920-8
- [52] Tanaka F, Umino M, Maeda M, Nakayama R, Inoue K, Kogue R, et al. Pseudocontinuous Arterial Spin Labeling: Clinical Applications and Usefulness in Head and Neck Entities. *Cancers*. 2022;14:3872
- [53] Langendijk JA, Hoebbers FJP, de Jong MA, Doornaert P, Terhaard CHJ, Steenbakkers RJHM, et al. National Protocol for Model-Based Selection for Proton Therapy in Head and Neck Cancer. *Int J Part Ther*. 2021;8:354-65
- [54] Moon DH, Moon SH, Wang K, Weissler MC, Hackman TG, Zanation AM, et al. Incidence of, and risk factors for, mandibular osteoradionecrosis in patients with oral cavity and oropharynx cancers. *Oral Oncol*. 2017;72:98-103
- [55] Aarup-Kristensen S, Hansen CR, Forner L, Brink C, Eriksen JG, Johansen J. Osteoradionecrosis of the mandible after radiotherapy for head and neck cancer: risk factors and dose-volume correlations. *Acta Oncol*. 2019;58:1373-7
- [56] Noij DP, Martens RM, Marcus JT, de Bree R, Leemans CR, Castelijns JA, et al. Intravoxel incoherent motion magnetic resonance imaging in head and neck cancer: A systematic review of the diagnostic and prognostic value. *Oral Oncol*. 2017;68:81-91
- [57] Fujima N, Sakashita T, Homma A, Yoshida D, Kudo K, Shirato H. Utility of a Hybrid IVIM-DKI Model to Predict the Development of Distant Metastasis in Head and Neck Squamous Cell Carcinoma Patients. *Magn Reson Med Sci*. 2018;17:21-7
- [58] Rosenkrantz AB, Padhani AR, Chenevert TL, Koh DM, De Keyzer F, Taouli B, et al. Body diffusion kurtosis imaging: Basic principles, applications, and considerations for clinical practice. *J Magn Reson Imaging*. 2015;42:1190-202
- [59] Leibfarth S, Winter RM, Lyng H, Zips D, Thorwarth D. Potentials and challenges of diffusion-weighted magnetic resonance imaging in radiotherapy. *Clin Transl Radiat Oncol*. 2018;13:29-37
- [60] Bito Y, Hirata S, Yamamoto E. Optimum Gradient Factors for Apparent Diffusion Coefficient Measurements. 3rd Annual Meeting of ISMRM. Nice, France 1995. p. 913.
- [61] Xing D, Papadakis NG, Huang CLH, Lee VM, Carpenter TA, Hall LD. Optimised Diffusion-Weighting for Measurement of Apparent Diffusion Coefficient (ADC) in Human Brain. *Magn Reson Imaging*. 1997;15:771-84
- [62] Brihuega-Moreno O, Heese FP, Hall LD. Optimization of diffusion measurements using Cramer-Rao lower bound theory and its application to articular cartilage. *Magn Reson Med*. 2003;50:1069-76

References

- [63] Peña-Nogales O, Hernando D, Aja-Fernandez S, de Luis-Garcia R. Determination of optimized set of b-values for Apparent Diffusion Coefficient mapping in liver Diffusion-Weighted MRI. *J Magn Reson*. 2020;310:106634
- [64] Kingsley PB, Monahan WG. Selection of the optimum b factor for diffusion-weighted magnetic resonance imaging assessment of ischemic stroke. *Magn Reson Med*. 2004;51:996-1001
- [65] Saritas EU, Lee JH, Nishimura DG. SNR dependence of optimal parameters for apparent diffusion coefficient measurements. *IEEE Trans Med Imaging*. 2011;30:424-37
- [66] Merisaari H, Jambor I. Optimization of b-value distribution for four mathematical models of prostate cancer diffusion-weighted imaging using b values up to 2000 s/mm²: simulation and repeatability study. *Magn Reson Med*. 2015;73:1954-69
- [67] Leporq B, Saint-Jalmes H, Rabrait C, Pilleul F, Guillaud O, Dumortier J, et al. Optimization of intra-voxel incoherent motion imaging at 3.0 Tesla for fast liver examination. *J Magn Reson Imaging*. 2015;41:1209-17
- [68] Zhang JL, Sigmund EE, Rusinek H, Chandarana H, Storey P, Chen Q, et al. Optimization of b-value sampling for diffusion-weighted imaging of the kidney. *Magn Reson Med*. 2012;67:89-97
- [69] Lemke A, Stieltjes B, Schad LR, Laun FB. Toward an optimal distribution of b values for intravoxel incoherent motion imaging. *Magn Reson Imaging*. 2011;29:766-76
- [70] Jalnefjord O, Montelius M, Starck G, Ljungberg M. Optimization of b-value schemes for estimation of the diffusion coefficient and the perfusion fraction with segmented intravoxel incoherent motion model fitting. *Magn Reson Med*. 2019;82:1541-52
- [71] Wang X, Hickernell FJ. Randomized Halton sequences. *Math Comput Model*. 2000;32:887-99
- [72] Lu Y, Jansen JFA, Stambuk HE, Gupta G, Lee N, Gonen M, et al. Comparing primary tumors and metastatic nodes in head and neck cancer using intravoxel incoherent motion imaging: a preliminary experience. *J Comput Assist Tomogr*. 2013;37:346-52
- [73] Paudyal R, Oh JH, Riaz N, Venigalla P, Li J, Hatzoglou V, et al. Intravoxel incoherent motion diffusion-weighted MRI during chemoradiation therapy to characterize and monitor treatment response in human papillomavirus head and neck squamous cell carcinoma. *J Magn Reson Imaging*. 2017;45:1013-23
- [74] Marzi S, Piludu F, Vidiri A. Assessment of diffusion parameters by intravoxel incoherent motion MRI in head and neck squamous cell carcinoma. *NMR Biomed*. 2013;26:1806-14
- [75] Alexander DC. A general framework for experiment design in diffusion MRI and its application in measuring direct tissue-microstructure features. *Magn Reson Med*. 2008;60:439-48
- [76] Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*. 2004;23 Suppl 1:S208-19
- [77] Guyader JM, Bernardin L, Douglas NHM, Poot DHJ, Niessen WJ, Klein S. Influence of image registration on apparent diffusion coefficient images computed from free-breathing diffusion MR images of the abdomen. *J Magn Reson Imaging*. 2015;42:315-30

- [78] Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2010;29:196-205
- [79] Bland JM, Altman DG. Measurement error. *BMJ*. 1996;313:744
- [80] Poot DHJ, den Dekker AJ, Achten E, Verhoye M, Sijbers J. Optimal experimental design for diffusion kurtosis imaging. *IEEE Trans Med Imaging*. 2010;29:819-29
- [81] Kolff-Gart AS, Pouwels PJW, Noij DP, Ljumanovic R, Vandecaveye V, de Keyzer F, et al. Diffusion-weighted imaging of the head and neck in healthy subjects: reproducibility of ADC values in different MRI systems and repeat sessions. *AJNR Am J Neuroradiol*. 2015;36:384-90
- [82] Chevallier O, Zhou N, He J, Loffroy R, Wang YXJ. Removal of evidential motion-contaminated and poorly fitted image data improves IVIM diffusion MRI parameter scan-rescan reproducibility. *Acta Radiologica*. 2018;59:1157-67
- [83] Gurney-Champion OJ, Froeling M, Klaassen R, Runge JH, Bel A, van Laarhoven HWM, et al. Minimizing the Acquisition Time for Intravoxel Incoherent Motion Magnetic Resonance Imaging Acquisitions in the Liver and Pancreas. *Invest Radiol*. 2016;51:211-20
- [84] Pai VM, Rapacchi S, Kellman P, Croisille P, Wen H. PCATMIP: enhancing signal intensity in diffusion-weighted magnetic resonance imaging. *Magn Reson Med*. 2011;65:1611-9
- [85] Manfrini E, Smits M, Thust S, Geiger S, Bendella Z, Petr J, et al. From research to clinical practice: a European neuroradiological survey on quantitative advanced MRI implementation. *Eur Radiol*. 2021;31:6334-41
- [86] Martin-Noguerol T, Kirsch CFE, Montesinos P, Luna A. Arterial spin labeling for head and neck lesion assessment: technical adjustments and clinical applications. *Neuroradiology*. 2021;63:1969-83
- [87] Kato H, Kanematsu M, Watanabe H, Kajita K, Mizuta K, Aoki M, et al. Perfusion imaging of parotid gland tumours: usefulness of arterial spin labeling for differentiating Warthin's tumours. *Eur Radiol*. 2015;25:3247-54
- [88] Yamamoto T, Kimura H, Hayashi K, Imamura Y, Mori M. Pseudo-continuous arterial spin labeling MR images in Warthin tumors and pleomorphic adenomas of the parotid gland: qualitative and quantitative analyses and their correlation with histopathologic and DWI and dynamic contrast enhanced MRI findings. *Neuroradiology*. 2018;60:803-12
- [89] Abdel Razek AAK, Nada N. Arterial spin labeling perfusion-weighted MR imaging: correlation of tumor blood flow with pathological degree of tumor differentiation, clinical stage and nodal metastasis of head and neck squamous cell carcinoma. *Eur Arch Otorhinolaryngol*. 2018;275:1301-7
- [90] Li Y, Li X, Yu X, Lin M, Ouyang H, Xie L, et al. Investigating the value of arterial spin labeling and intravoxel incoherent motion imaging on diagnosing nasopharyngeal carcinoma in T1 stage. *Cancer Imaging*. 2020;20:62
- [91] Abdel Razek AAK, Helmy E. Multi-parametric arterial spin labeling and diffusion-weighted imaging in differentiation of metastatic from reactive lymph nodes in head and neck squamous cell carcinoma. *Eur Arch Otorhinolaryngol*. 2021;278:2529-35
- [92] Fujima N, Yoshida D, Sakashita T, Homma A, Tsukahara A, Tha KK, et al. Usefulness of Pseudocontinuous Arterial Spin-Labeling for the Assessment of Patients with Head and Neck Squamous Cell Carcinoma by Measuring Tumor Blood Flow in the Pretreatment and Early Treatment Period. *AJNR Am J Neuroradiol*. 2016;37:342-8

References

- [93] Fujima N, Kudo K, Tsukahara A, Yoshida D, Sakashita T, Homma A, et al. Measurement of tumor blood flow in head and neck squamous cell carcinoma by pseudo-continuous arterial spin labeling: comparison with dynamic contrast-enhanced MRI. *J Magn Reson Imaging*. 2015;41:983-91
- [94] Lin M, Yu X, Luo D, Ouyang H, Xie L, Wu B, et al. Investigating the correlation of arterial spin labeling and dynamic contrast enhanced perfusion in primary tumor of nasopharyngeal carcinoma. *Eur J Radiol*. 2018;108:222-9
- [95] Houweling AC, Schakel T, van den Berg CAT, Philippens MEP, Roesink JM, Terhaard CHJ, et al. MRI to quantify early radiation-induced changes in the salivary glands. *Radiother Oncol*. 2011;100:386-9
- [96] Lee FK, King AD, Kam MK, Ma BB, Yeung DK. Radiation injury of the parotid glands during treatment for head and neck cancer: assessment using dynamic contrast-enhanced MR imaging. *Radiat Res*. 2011;175:291-6
- [97] Fujita S, Hagiwara A, Hori M, Warntjes M, Kamagata K, Fukunaga I, et al. Three-dimensional high-resolution simultaneous quantitative mapping of the whole brain with 3D-QALAS: An accuracy and repeatability study. *Magn Reson Imaging*. 2019;63:235-43
- [98] Dai W, Robson PM, Shankaranarayanan A, Alsop DC. Reduced resolution transit delay prescan for quantitative continuous arterial spin labeling perfusion imaging. *Magn Reson Med*. 2012;67:1252-65
- [99] Dai W, Shankaranarayanan A, Alsop DC. Volumetric measurement of perfusion and arterial transit delay using hadamard encoded continuous arterial spin labeling. *Magn Reson Med*. 2013;69:1014-22
- [100] Maleki N, Dai W, Alsop DC. Optimization of background suppression for arterial spin labeling perfusion imaging. *MAGMA*. 2012;25:127-33
- [101] Lu H, Clingman C, Golay X, van Zijl PCM. Determining the longitudinal relaxation time (T1) of blood at 3.0 Tesla. *Magn Reson Med*. 2004;52:679-82
- [102] Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. 2015;67:1-48
- [103] Barnhart HX, Barboriak DP. Applications of the repeatability of quantitative imaging biomarkers: a review of statistical analysis of repeat data sets. *Transl Oncol*. 2009;2:231-5
- [104] Hales PW, Kawadler JM, Aylett SE, Kirkham FJ, Clark CA. Arterial spin labeling characterization of cerebral perfusion during normal maturation from late childhood into adulthood: normal 'reference range' values and their use in clinical studies. *J Cereb Blood Flow Metab*. 2014;34:776-84
- [105] Zavala Bojorquez J, Bricq S, Acquitter C, Brunotte F, Walker PM, Lalande A. What are normal relaxation times of tissues at 3 T? *Magn Reson Imaging*. 2017;35:69-80
- [106] Abdel Razek AAK. Multi-parametric MR imaging using pseudo-continuous arterial-spin labeling and diffusion-weighted MR imaging in differentiating subtypes of parotid tumors. *Magn Reson Imaging*. 2019;63:55-9
- [107] Lin M, Yu X, Ouyang H, Luo D, Zhou C. Consistency of T2WI-FS/ASL fusion images in delineating the volume of nasopharyngeal carcinoma. *Sci Rep*. 2015;5:18431
- [108] Chappell MA, MacIntosh BJ, Donahue MJ, Gunther M, Jezzard P, Woolrich MW. Separation of macrovascular signal in multi-inversion time arterial spin labelling MRI. *Magn Reson Med*. 2010;63:1357-65
- [109] Wheeler RH, Ziessman HA, Medvec BR, Juni JE, Thrall JH, Keyes JW, et al. Tumor Blood Flow and Systemic Shunting in Patients Receiving Intraarterial Chemotherapy for Head and Neck Cancer. *Cancer Res*. 1986;46:4200-04

- [110] Herscovitch P, Raichle ME. What Is the Correct Value for the Brain-Blood Partition Coefficient for Water? *Journal of Cerebral Blood Flow and Metabolism*. 1985;5:65-9
- [111] Hirata K, Hattori N, Katoh C, Shiga T, Kuroda S, Kubo N, et al. Regional partition coefficient of water in patients with cerebrovascular disease and its effect on rCBF assessment. *Nucl Med Commun*. 2011;32:63-70
- [112] Federau C. Intravoxel incoherent motion MRI as a means to measure in vivo perfusion: A review of the evidence. *NMR Biomed*. 2017;30
- [113] Le Bihan D, Turner R. The capillary network: a link between IVIM and classical perfusion. *Magn Reson Med*. 1992;27:171-8
- [114] Ma G, Xu XQ, Zhu LN, Jiang JS, Su GY, Hu H, et al. Intravoxel Incoherent Motion Magnetic Resonance Imaging for Assessing Parotid Gland Tumors: Correlation and Comparison with Arterial Spin Labeling Imaging. *Korean J Radiol*. 2021;22:243-52
- [115] Wu W, Jiang G, Xu Z, Wang R, Pan A, Gao M, et al. Three-dimensional pulsed continuous arterial spin labeling and intravoxel incoherent motion imaging of nasopharyngeal carcinoma: correlations with Ki-67 proliferation status. *Quant Imaging Med Surg*. 2021;11:1394-405
- [116] Dolgosuren EA, Harada M, Kanazawa Y, Abe T, Otomo M, Matsumoto Y, et al. Correlation and Characteristics of Intravoxel Incoherent Motion and Arterial Spin Labeling Techniques Versus Multiple Parameters Obtained on Dynamic Susceptibility Contrast Perfusion MRI for Brain Tumors. *The Journal of Medical Investigation*. 2019;66:308-13
- [117] Shen N, Zhao L, Jiang J, Jiang R, Su C, Zhang S, et al. Intravoxel incoherent motion diffusion-weighted imaging analysis of diffusion and microperfusion in grading gliomas and comparison with arterial spin labeling for evaluation of tumor perfusion. *J Magn Reson Imaging*. 2016;44:620-32
- [118] Stieb S, Boss A, Wurnig MC, Ozbay PS, Weiss T, Guckenberger M, et al. Non-parametric intravoxel incoherent motion analysis in patients with intracranial lesions: Test-retest reliability and correlation with arterial spin labeling. *Neuroimage Clin*. 2016;11:780-8
- [119] Lin Y, Li J, Zhang Z, Xu Q, Zhou Z, Zhang Z, et al. Comparison of Intravoxel Incoherent Motion Diffusion-Weighted MR Imaging and Arterial Spin Labeling MR Imaging in Gliomas. *Biomed Res Int*. 2015;2015:234245
- [120] Sijtsma ND, Petit SF, Verduijn GM, Poot DHJ, Warnert EAH, Hoogeman MS, et al. Multidelay pseudocontinuous arterial spin labeling to measure blood flow in the head and neck. *NMR Biomed*. 2023;36:e4898
- [121] Sijtsma ND, Petit SF, Poot DHJ, Verduijn GM, van der Lugt A, Hoogeman MS, et al. An optimal acquisition and post-processing pipeline for hybrid IVIM-DKI in head and neck. *Magn Reson Med*. 2021;85:777-89
- [122] R Core Team. R: A language and environment for statistical computing. <https://www.R-project.org/>
- [123] Plummer M. rjags: Bayesian Graphical Models using MCMC. <https://CRAN.R-project.org/package=rjags>
- [124] Le Bihan D. What can we see with IVIM MRI? *Neuroimage*. 2019;187:56-67
- [125] Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer*. 2010;127:2893-917
- [126] Blanchard P, Baujat B, Holostenco V, Bourredjem A, Baey C, Bourhis J, et al. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): a comprehensive analysis by tumour site. *Radiother Oncol*. 2011;100:33-40

References

- [127] Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tân PF, et al. Human Papillomavirus and Survival of Patients with Oropharyngeal Cancer. *N Engl J Med.* 2010;363:24-35
- [128] Wan Leung S, Lee TF, Chien CY, Chao PJ, Tsai WL, Fang FM. Health-related quality of life in 640 head and neck cancer survivors after radiotherapy using EORTC QLQ-C30 and QLQ-H&N35 questionnaires. *BMC Cancer.* 2011;11
- [129] Hunter KU, Schipper M, Feng FY, Lyden T, Haxer M, Murdoch-Kinch CA, et al. Toxicities affecting quality of life after chemo-IMRT of oropharyngeal cancer: prospective study of patient-reported, observer-rated, and objective outcomes. *Int J Radiat Oncol Biol Phys.* 2013;85:935-40
- [130] Verdonck-de Leeuw IM, Buffart LM, Heymans MW, Rietveld DH, Doornaert P, de Bree R, et al. The course of health-related quality of life in head and neck cancer patients treated with chemoradiation: a prospective cohort study. *Radiother Oncol.* 2014;110:422-8
- [131] Van den Bosch L, van der Laan HP, van der Schaaf A, Oosting SF, Halmos GB, Witjes MJH, et al. Patient-Reported Toxicity and Quality-of-Life Profiles in Patients With Head and Neck Cancer Treated With Definitive Radiation Therapy or Chemoradiation. *Int J Radiat Oncol Biol Phys.* 2021;111:456-67
- [132] Linge A, Lohaus F, Lock S, Nowak A, Gudziol V, Valentini C, et al. HPV status, cancer stem cell marker expression, hypoxia gene signatures and tumour volume identify good prognosis subgroups in patients with HNSCC after primary radiochemotherapy: A multicentre retrospective study of the German Cancer Consortium Radiation Oncology Group (DKTK-ROG). *Radiother Oncol.* 2016;121:364-73
- [133] Knegjens JL, Hauptmann M, Pameijer FA, Balm AJ, Hoebbers FJ, de Bois JA, et al. Tumor volume as prognostic factor in chemoradiation for advanced head and neck cancer. *Head Neck.* 2011;33:375-82
- [134] Lassen P, Lacas B, Pignon JP, Trotti A, Zackrisson B, Zhang Q, et al. Prognostic impact of HPV-associated p16-expression and smoking status on outcomes following radiotherapy for oropharyngeal cancer: The MARCH-HPV project. *Radiother Oncol.* 2018;126:107-15
- [135] Rietbergen MM, Brakenhoff RH, Bloemena E, Witte BI, Snijders PJF, Heideman DAM, et al. Human papillomavirus detection and comorbidity: critical issues in selection of patients with oropharyngeal cancer for treatment De-escalation trials. *Ann Oncol.* 2013;24:2740-5
- [136] Rietbergen MM, Witte BI, Velazquez ER, Snijders PJF, Bloemena E, Speel EJ, et al. Different prognostic models for different patient populations: validation of a new prognostic model for patients with oropharyngeal cancer in Western Europe. *Br J Cancer.* 2015;112:1733-6
- [137] van der Schroeff MP, Steyerberg EW, Wieringa MH, Langeveld TPM, Molenaar J, Baatenburg de Jong RJ. Prognosis: a variable parameter: dynamic prognostic modeling in head and neck squamous cell carcinoma. *Head Neck.* 2012;34:34-41
- [138] Leemans CR, Snijders PJF, Brakenhoff RH. The molecular landscape of head and neck cancer. *Nat Rev Cancer.* 2018;18:269-82
- [139] Naipal KAT, Verkaik NS, Sanchez H, van Deurzen CHM, den Bakker MA, Hoeijmakers JHJ, et al. Tumor slice culture system to assess drug response of primary breast cancer. *BMC Cancer.* 2016;16:78
- [140] Goldberg SB, Narayan A, Kole AJ, Decker RH, Teysir J, Carriero NJ, et al. Early Assessment of Lung Cancer Immunotherapy Response via Circulating Tumor DNA. *Clin Cancer Res.* 2018;24:1872-80

- [141] Cheng H, Liu C, Jiang J, Luo G, Lu Y, Jin K, et al. Analysis of ctDNA to predict prognosis and monitor treatment responses in metastatic pancreatic cancer patients. *Int J Cancer*. 2017;140:2344–50
- [142] van Ginkel JH, Sliker FJB, de Bree R, van Es RJJ, Van Cann EM, Willems SM. Cell-free nucleic acids in body fluids as biomarkers for the prediction and early detection of recurrent head and neck cancer: A systematic review of the literature. *Oral Oncol*. 2017;75:8–15
- [143] Wang Y, Springer S, Mulvey CL, Silliman N, Schaefer J, Sausen M, et al. Detection of somatic mutations and HPV in the saliva and plasma of patients with head and neck squamous cell carcinomas. *Sci Transl Med*. 2015;7:293ra104
- [144] Capala ME, Verduijn GM, Petit SF, de Korte MA, Hardillo JA, Sewnaik A, et al. Ex vivo functional assay for predicting radiation treatment response in squamous cell carcinoma of the head and neck. *Oral Oncol*. 2021;118:4–5
- [145] Burgener JM, Zou J, Zhao Z, Zheng Y, Shen SY, Huang SH, et al. Tumor-Naïve Multimodal Profiling of Circulating Tumor DNA in Head and Neck Squamous Cell Carcinoma. *Clin Cancer Res*. 2021;27:4230–44
- [146] Lok BH, Setton J, Caria N, Romanyshyn J, Wolden SL, Zelefsky MJ, et al. Intensity-modulated radiation therapy in oropharyngeal carcinoma: effect of tumor volume on clinical outcomes. *Int J Radiat Oncol Biol Phys*. 2012;82:1851–7
- [147] Carpén T, Saarilahti K, Haglund C, Markkola A, Tarkkanen J, Hagstrom J, et al. Tumor volume as a prognostic marker in p16-positive and p16-negative oropharyngeal cancer patients treated with definitive intensity-modulated radiotherapy. *Strahlenther Onkol*. 2018;194:759–70
- [148] Studer G, Lutolf UM, El-Bassiouni M, Rousson V, Glanzmann C. Volumetric staging (VS) is superior to TNM and AJCC staging in predicting outcome of head and neck cancer treated with IMRT. *Acta Oncol*. 2007;46:386–94
- [149] Lambrecht M, Van Calster B, Vandecaveye V, De Keyzer F, Roebben I, Hermans R, et al. Integrating pretreatment diffusion weighted MRI into a multivariable prognostic model for head and neck squamous cell carcinoma. *Radiother Oncol*. 2014;110:429–34
- [150] King AD, Thoeny HC. Functional MRI for the prediction of treatment response in head and neck squamous cell carcinoma: potential and limitations. *Cancer Imaging*. 2016;16:23
- [151] Ding Y, Hazle JD, Mohamed ASR, Frank SJ, Hobbs BP, Colen RR, et al. Intravoxel incoherent motion imaging kinetics during chemoradiotherapy for human papillomavirus-associated squamous cell carcinoma of the oropharynx: preliminary results from a prospective pilot study. *NMR Biomed*. 2015;28:1645–54
- [152] Hentschel M, Appold S, Schreiber A, Abolmaali N, Abramyuk A, Dorr W, et al. Early FDG PET at 10 or 20 Gy under chemoradiotherapy is prognostic for locoregional control and overall survival in patients with head and neck cancer. *Eur J Nucl Med Mol Imaging*. 2011;38:1203–11
- [153] Kulasinghe A, Perry C, Jovanovic L, Nelson C, Punyadeera C. Circulating tumour cells in metastatic head and neck cancers. *Int J Cancer*. 2015;136:2515–23
- [154] Ravanelli M, Grammatica A, Maddalo M, Ramanzin M, Agazzi GM, Tononcelli E, et al. Pretreatment DWI with Histogram Analysis of the ADC in Predicting the Outcome of Advanced Oropharyngeal Cancer with Known Human Papillomavirus Status Treated with Chemoradiation. *AJNR Am J Neuroradiol*. 2020;41:1473–9

References

- [155] Martens RM, Koopman T, Lavini C, Ali M, Peeters CFW, Noij DP, et al. Multiparametric functional MRI and (18)F-FDG-PET for survival prediction in patients with head and neck squamous cell carcinoma treated with (chemo) radiation. *Eur Radiol.* 2021;31:616–28
- [156] Freihat O, Toth Z, Pinter T, Kedves A, Sipos D, Cselik Z, et al. Pre-treatment PET/MRI based FDG and DWI imaging parameters for predicting HPV status and tumor response to chemoradiotherapy in primary oropharyngeal squamous cell carcinoma (OPSCC). *Oral Oncol.* 2021;116:105239
- [157] Sistonen HJ, Aro K, Atula T, Jouhi L, Linden R, Tapiovaara L, et al. Prognostic Value of Apparent Diffusion Coefficient in Oropharyngeal Carcinoma. *Clin Neuroradiol.* 2021;31:1037–48
- [158] Ravanelli M, Grammatica A, Tononcelli E, Morello R, Leali M, Battocchio S, et al. Correlation between Human Papillomavirus Status and Quantitative MR Imaging Parameters including Diffusion-Weighted Imaging and Texture Features in Oropharyngeal Carcinoma. *AJNR Am J Neuroradiol.* 2018;39:1878–83
- [159] Reeder SB, Pineda AR, Wen Z, Shimakawa A, Yu H, Brittain JH, et al. Iterative decomposition of water and fat with echo asymmetry and least-squares estimation (IDEAL): application with fast spin-echo imaging. *Magn Reson Med.* 2005;54:636–44
- [160] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Transactions on pattern analysis and machine intelligence.* 2012;34:2274–81
- [161] Ward JH. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association.* 1963;58:236–44
- [162] Even AJG, Reymen B, La Fontaine MD, Das M, Mottaghy FM, Belderbos JSA, et al. Clustering of multi-parametric functional imaging to identify high-risk subvolumes in non-small cell lung cancer. *Radiother Oncol.* 2017;125:379–84
- [163] Caliński T, Harabasz J. A dendrite method for cluster analysis. *Communications in Statistics.* 1974;3:1–27
- [164] Bishop JA, Lewis JSJ, Rocco JW, Faquin WC. HPV-related squamous cell carcinoma of the head and neck: An update on testing in routine pathology practice. *Semin Diagn Pathol.* 2015;32:344–51
- [165] El-Naggar AK, Westra WH. p16 expression as a surrogate marker for HPV-related oropharyngeal carcinoma: a guide for interpretative relevance and consistency. *Head Neck.* 2012;34:459–61
- [166] Liu SZ, Zandberg DP, Schumaker LM, Papadimitriou JC, Cullen KJ. Correlation of p16 expression and HPV type with survival in oropharyngeal squamous cell cancer. *Oral Oncol.* 2015;51:862–9
- [167] Piludu F, Marzi S, Gangemi E, Farneti A, Marucci L, Venuti A, et al. Multiparametric MRI Evaluation of Oropharyngeal Squamous Cell Carcinoma. A Mono-Institutional Study. *J Clin Med.* 2021;10
- [168] Han M, Lee SJ, Lee D, Kim SY, Choi JW. Correlation of human papilloma virus status with quantitative perfusion/diffusion/metabolic imaging parameters in the oral cavity and oropharyngeal squamous cell carcinoma: comparison of primary tumour sites and metastatic lymph nodes. *Clin Radiol.* 2018;73:757.e21–.e27
- [169] Lenoir V, Delattre BMA, M’Ra DY, De Vito C, de Perrot T, Becker M. Diffusion-Weighted Imaging to Assess HPV-Positive versus HPV-Negative Oropharyngeal Squamous Cell Carcinoma: The Importance of b-Values. *AJNR Am J Neuroradiol.* 2022;43:905–12

- [170] Driessen JP, van Bommel AJM, van Kempen PMW, Janssen LM, Terhaard CHJ, Pameijer FA, et al. Correlation of human papillomavirus status with apparent diffusion coefficient of diffusion-weighted MRI in head and neck squamous cell carcinomas. *Head Neck*. 2016;38 Suppl 1:E613-8
- [171] Nakahira M, Saito N, Yamaguchi H, Kuba K, Sugasawa M. Use of quantitative diffusion-weighted magnetic resonance imaging to predict human papilloma virus status in patients with oropharyngeal squamous cell carcinoma. *Eur Arch Otorhinolaryngol*. 2014;271:1219-25
- [172] Peltenburg B, Driessen JP, Vasmel JE, Pameijer FA, Janssen LM, Terhaard CHJ, et al. Pretreatment ADC is not a prognostic factor for local recurrences in head and neck squamous cell carcinoma when clinical T-stage is known. *Eur Radiol*. 2020;30:1228-31
- [173] Vidiri A, Marzi S, Gangemi E, Benevolo M, Rollo F, Farneti A, et al. Intravoxel incoherent motion diffusion-weighted imaging for oropharyngeal squamous cell carcinoma: Correlation with human papillomavirus Status. *Eur J Radiol*. 2019;119:108640
- [174] Cao Y, Aryal M, Li P, Lee C, Schipper M, Hawkins PG, et al. Predictive Values of MRI and PET Derived Quantitative Parameters for Patterns of Failure in Both p16+ and p16- High Risk Head and Neck Cancer. *Front Oncol*. 2019;9:1118
- [175] Chan MW, Higgins K, Enepekides D, Poon I, Symons SP, Moineddin R, et al. Radiologic Differences between Human Papillomavirus-Related and Human Papillomavirus-Unrelated Oropharyngeal Carcinoma on Diffusion-Weighted Imaging. *ORL J Otorhinolaryngol Relat Spec*. 2016;78:344-52
- [176] Connor S, Anjari M, Burd C, Guha A, Lei M, Guerrero-Urbano T, et al. The impact of Human Papilloma Virus status on the prediction of head and neck cancer chemoradiotherapy outcomes using the pre-treatment apparent diffusion coefficient. *Br J Radiol*. 2022;95:20210333
- [177] Schouten CS, de Graaf P, Bloemena E, Witte BI, Braakhuis BJM, Brakenhoff RH, et al. Quantitative diffusion-weighted MRI parameters and human papillomavirus status in oropharyngeal squamous cell carcinoma. *AJNR Am J Neuroradiol*. 2015;36:763-7
- [178] Hatakenaka M, Nakamura K, Yabuuchi H, Shioyama Y, Matsuo Y, Ohnishi K, et al. Pretreatment apparent diffusion coefficient of the primary lesion correlates with local failure in head-and-neck cancer treated with chemoradiotherapy or radiotherapy. *Int J Radiat Oncol Biol Phys*. 2011;81:339-45
- [179] Chawla S, Kim S, Dougherty L, Wang S, Loevner LA, Quon H, et al. Pretreatment diffusion-weighted and dynamic contrast-enhanced MRI for prediction of local treatment response in squamous cell carcinomas of the head and neck. *AJR Am J Roentgenol*. 2013;200:35-43
- [180] Thompson LDR, Burchette R, Iganej S, Bhattasali O. Oropharyngeal Squamous Cell Carcinoma in 390 Patients: Analysis of Clinical and Histological Criteria Which Significantly Impact Outcome. *Head Neck Pathol*. 2020;14:666-88
- [181] Hanns E, Job S, Coliat P, Wasyluk C, Ramolu L, Pencreach E, et al. Human Papillomavirus-related tumours of the oropharynx display a lower tumour hypoxia signature. *Oral Oncol*. 2015;51:848-56
- [182] Mortensen LS, Johansen J, Kallehauge J, Primdahl H, Busk M, Lassen P, et al. FAZA PET/CT hypoxia imaging in patients with squamous cell carcinoma of the head and neck treated with radiotherapy: results from the DAHANCA 24 trial. *Radiother Oncol*. 2012;105:14-20

References

- [183] Matoba M, Tuji H, Shimode Y, Toyoda I, Kuginuki Y, Miwa K, et al. Fractional change in apparent diffusion coefficient as an imaging biomarker for predicting treatment response in head and neck cancer treated with chemoradiotherapy. *AJNR Am J Neuroradiol.* 2014;35:379-85
- [184] Hauser T, Essig M, Jensen A, Laun FB, Munter M, Maier-Hein KH, et al. Prediction of lymph node metastasis. *Eur J Radiol.* 2014;83:783-7
- [185] Wang H, Zhang Y, Bai W, Wang B, Wei J, Ji R, et al. Feasibility of Immunohistochemical p16 Staining in the Diagnosis of Human Papillomavirus Infection in Patients With Squamous Cell Carcinoma of the Head and Neck: A Systematic Review and Meta-Analysis. *Front Oncol.* 2020;10:524928
- [186] Støre G, Boysen M. Mandibular osteoradionecrosis: clinical behaviour and diagnostic aspects. *Clin Otolaryngol.* 2000;25:378-84
- [187] Jereczek-Fossa BA, Orecchia R. Radiotherapy-induced mandibular bone complications. *Cancer Treat Rev.* 2002;28:65-74
- [188] Schwartz HC, Kagan AR. Osteoradionecrosis of the mandible: scientific basis for clinical staging. *Am J Clin Oncol.* 2002;25:168-71
- [189] Mohamed ASR, Hobbs BP, Hutcheson KA, Murri MS, Garg N, Song J, et al. Dose-volume correlates of mandibular osteoradionecrosis in Oropharynx cancer patients receiving intensity-modulated radiotherapy: Results from a case-matched comparison. *Radiother Oncol.* 2017;124:232-9
- [190] De Maesschalck T, Dulguerov N, Caparrotti F, Scolozzi P, Picardi C, Mach N, et al. Comparison of the incidence of osteoradionecrosis with conventional radiotherapy and intensity-modulated radiotherapy. *Head Neck.* 2016;38:1695-702
- [191] Zhang W, Zhang X, Yang P, Blanchard P, Garden AS, Gunn B, et al. Intensity-modulated proton therapy and osteoradionecrosis in oropharyngeal cancer. *Radiother Oncol.* 2017;123:401-5
- [192] van Dijk LV, Abusaif AA, Rigert J, Naser MA, Hutcheson KA, Lai SY, et al. Normal Tissue Complication Probability (NTCP) Prediction Model for Osteoradionecrosis of the Mandible in Patients With Head and Neck Cancer After Radiation Therapy: Large-Scale Observational Cohort. *Int J Radiat Oncol Biol Phys.* 2021;111:549-58
- [193] Tsai CJ, Hofstede TM, Sturgis EM, Garden AS, Lindberg ME, Wei Q, et al. Osteoradionecrosis and radiation dose to the mandible in patients with oropharyngeal cancer. *Int J Radiat Oncol Biol Phys.* 2013;85:415-20
- [194] Caparrotti F, Huang SH, Lu L, Bratman SV, Ringash J, Bayley A, et al. Osteoradionecrosis of the mandible in patients with oropharyngeal carcinoma treated with intensity-modulated radiotherapy. *Cancer.* 2017;123:3691-700
- [195] Owosho AA, Tsai CJ, Lee RS, Freymiller H, Kadempour A, Varthi S, et al. The prevalence and risk factors associated with osteoradionecrosis of the jaw in oral and oropharyngeal cancer patients treated with intensity-modulated radiation therapy (IMRT): The Memorial Sloan Kettering Cancer Center experience. *Oral Oncol.* 2017;64:44-51
- [196] Bedwinek JM, Shukovsky LJ, Fletcher GH, Daley TE. Osteonecrosis in patients treated with definitive radiotherapy for squamous cell carcinomas of the oral cavity and naso-and oropharynx. *Radiology.* 1976;119:665-7

- [197] Ben-David MA, Diamante M, Radawski JD, Vineberg KA, Stroup C, Murdoch-Kinch CA, et al. Lack of osteoradionecrosis of the mandible after intensity-modulated radiotherapy for head and neck cancer: likely contributions of both dental care and improved dose distributions. *Int J Radiat Oncol Biol Phys.* 2007;68:396-402
- [198] Lee CT, Litwin S, Yao CMKL, Liu JC, Ridge JA, Galloway TJ. Osteoradionecrosis rate in oropharynx cancer treated with dose volume histogram based constraints. *Radiother Oncol.* 2022;176:215-21
- [199] Al-Mamgani A, Kwa SLS, Tans L, Moring M, Fransen D, Mehilal R, et al. Single Vocal Cord Irradiation: Image Guided Intensity Modulated Hypofractionated Radiation Therapy for T1a Glottic Cancer: Early Clinical Results. *Int J Radiat Oncol Biol Phys.* 2015;93:337-43
- [200] Shuryak I, Hall EJ, Brenner DJ. Optimized Hypofractionation Can Markedly Improve Tumor Control and Decrease Late Effects for Head and Neck Cancer. *Int J Radiat Oncol Biol Phys.* 2019;104:272-8
- [201] Elbers JBW, Gunsch PA, Debets R, Keereweer S, van Meerten E, Zindler J, et al. Hypofractionated, dose-redistributed RAdiotherapy with protons and photons to combat radiation-induced immunosuppression in head and neck squamous cell carcinoma: study protocol of the phase I HYDRA trial. *BMC Cancer.* 2023;23:541
- [202] Withers HR, Peters LJ, Taylor JMG, Owen JB, Morrison WH, Schultheiss TE, et al. Late normal tissue sequelae from radiation therapy for carcinoma of the tonsil: patters of fractionation study of radiobiology. *Int J Radiat Oncol Biol Phys.* 1995;33:563-8
- [203] NVMKA. Richtlijn Mondholte- en Orofarynxcarcinoom. Nederlandse Vereniging voor Mondziekten, Kaak- en Aangezichts chirurgie. <https://www.nvmka.nl/files/Richtlijn-Mondholte--en-Orofarynxcarcinoom.pdf>
- [204] Kielbassa AM, Hinkelbein W, Hellwig E, Meyer-Luckel H. Radiation-related damage to dentition. *Lancet Oncol.* 2006;7:326-35
- [205] Niemierko A. Reporting and analyzing dose distributions A concept of equivalent uniform dose. *Medical Physics.* 1997;24:103-10
- [206] Wu Q, Mohan R, Niemierko A, Schmidt-Ullrich R. Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose. *Int J Radiat Oncol Biol Phys.* 2002;52:224-35
- [207] Wu Q, Djajaputra D, Liu HH, Dong L, Mohan R, Wu Y. Dose sculpting with generalized equivalent uniform dose. *Med Phys.* 2005;32:1387-96
- [208] Park CS, Kim Y, Lee N, Bucci KM, Quivey JM, Verhey LJ, et al. Method to account for dose fractionation in analysis of IMRT plans: modified equivalent uniform dose. *Int J Radiat Oncol Biol Phys.* 2005;62:925-32
- [209] Pyshniak V, Fotina I, Zverava A, Siamkouski S, Zayats E, Kopanitsa G, et al. Efficiency of biological versus physical optimization for single-arc VMAT for prostate and head and neck cases. *J Applied Clin Med Phys.* 2014;15:39-53
- [210] Can S, Ozer EE, Karacetin D. Various cost functions evaluation of commercial biologically based treatment planning system for nasopharyngeal cancer. *Med Dosim.* 2022;47:184-90
- [211] Chang DT, Sandow PR, Morris CG, Hollander R, Scarborough L, Amdur RJ, et al. Do pre-irradiation dental extractions reduce the risk of osteoradionecrosis of the mandible? *Head Neck.* 2007;29:528-36
- [212] Wang TH, Liu CJ, Chao TF, Chen TJ, Hu YW. Risk factors for and the role of dental extractions in osteoradionecrosis of the jaws: A national-based cohort study. *Head Neck.* 2017;39:1313-21

References

- [213] Sijtsema ND, Verduijn GM, Nasserinejad K, van Norden Y, Mast H, van der Lugt A, et al. Development of a local dose–response relationship for osteoradionecrosis within the mandible. *Radiother Oncol.* 2023;186:109736
- [214] Al-Mamgani A, Tans L, van Rooij PHE, Noever I, Baatenburg de Jong RJ, Levendag PC. Hypofractionated radiotherapy denoted as the “Christie scheme”: an effective means of palliating patients with head and neck cancers not suitable for curative treatment. *Acta Oncol.* 2009;48:562-70
- [215] Vargo JA, Ferris RL, Clump DA, Heron DE. Stereotactic body radiotherapy as primary treatment for elderly patients with medically inoperable head and neck cancer. *Front Oncol.* 2014;4:214
- [216] Laursen M, Specht L, Kristensen CA, Gothelf A, Bernsdorf M, Vogelius I, et al. An Extended Hypofractionated Palliative Radiotherapy Regimen for Head and Neck Carcinomas. *Front Oncol.* 2018;8:206
- [217] Eisbruch A, Harris J, Garden AS, Chao CK, Straube W, Harari PM, et al. Multi-institutional trial of accelerated hypofractionated intensity-modulated radiation therapy for early-stage oropharyngeal cancer (RTOG 00-22). *Int J Radiat Oncol Biol Phys.* 2010;76:1333–8
- [218] Lee DS, Kim YS, Cheon JS, Song JH, Son SH, Jang JS, et al. Long-term outcome and toxicity of hypofractionated stereotactic body radiotherapy as a boost treatment for head and neck cancer: the importance of boost volume assessment. *Radiat Oncol.* 2012;7
- [219] Karam SD, Rashid A, Snider JW, Wooster M, Bhatia S, Jay AK, et al. IMRT with Stereotactic Body Radiotherapy Boost for High Risk Malignant Salivary Gland Malignancies: A Case Series. *Front Oncol.* 2014;4:268
- [220] Baker S, Verduijn GM, Petit S, Sewnaik A, Mast H, Koljenovic S, et al. Long-term outcomes following stereotactic body radiotherapy boost for oropharyngeal squamous cell carcinoma. *Acta Oncol.* 2019;58:926–33
- [221] Kim MS, Malik NH, Chen H, Poon I, Husain Z, Eskander A, et al. Stereotactic radiotherapy as planned boost after definitive radiotherapy for head and neck cancers: Systematic review. *Head Neck.* 2022;44:770–82
- [222] Bras J, de Jonge HKT, van Merkesteyn JPR. Osteoradionecrosis of the Mandible: Pathogenesis. *Am J Otolaryngol.* 1990;11:244-50
- [223] Balermipas P, van Timmeren JE, Knierim DJ, Guckenberger M, Ciernik IF. Dental extraction, intensity-modulated radiotherapy of head and neck cancer, and osteoradionecrosis : A systematic review and meta-analysis. *Strahlenther Onkol.* 2022
- [224] Beaumont S, Bhatia N, McDowell L, Fua T, McCullough M, Celentano A, et al. Timing of dental extractions in patients undergoing radiotherapy and the incidence of osteoradionecrosis: a systematic review and meta-analysis. *Br J Oral Maxillofac Surg.* 2021;59:511-23
- [225] Focus onderzoek. Dutch Association of Oral and Maxillofacial Surgery NVMKA; 2021.
- [226] Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging.* 2004;23:903-21
- [227] Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika.* 1986;73:13-22
- [228] Zeger SL, Liang KY. Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics.* 1986;43:121-30

- [229] Kubota H, Miyawaki D, Mukumoto N, Ishihara T, Matsumura M, Hasegawa T, et al. Risk factors for osteoradionecrosis of the jaw in patients with head and neck squamous cell carcinoma. *Radiat Oncol.* 2021;16:1
- [230] Gomez DR, Estilo CL, Wolden SL, Zelefsky MJ, Kraus DH, Wong RJ, et al. Correlation of Osteoradionecrosis and Dental Events With Dosimetric Parameters in Intensity-Modulated Radiation Therapy for Head-and-Neck Cancer. *Int J Radiat Oncol Biol Phys.* 2011;81:e207-13
- [231] O'Connor JPB, Aboagye EO, Adams JE, Aerts HJWL, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol.* 2017;14:169-86
- [232] Meeus EM, Novak J, Withey SB, Zarinabad N, Dehghani H, Peet AC. Evaluation of intravoxel incoherent motion fitting methods in low-perfused tissue. *J Magn Reson Imaging.* 2017;45:1325-34
- [233] Suo S, Lin N, Wang H, Zhang L, Wang R, Zhang S, et al. Intravoxel incoherent motion diffusion-weighted MR imaging of breast cancer at 3.0 tesla: Comparison of different curve-fitting methods. *J Magn Reson Imaging.* 2015;42:362-70
- [234] Barbieri S, Gurney-Champion OJ, Klaassen R, Thoeny HC. Deep learning how to fit an intravoxel incoherent motion model to diffusion-weighted MRI. *Magn Reson Med.* 2020;83:312-21
- [235] Capala ME, Pachler KS, Lauwers I, de Korte MA, Verkaik NS, Mast H, et al. Ex Vivo Functional Assay for Evaluating Treatment Response in Tumor Tissue of Head and Neck Squamous Cell Carcinoma. *Cancers.* 2023;15:478
- [236] Lagendijk JJW, van Vulpen M, Raaymakers BW. The development of the MRI linac system for online MRI-guided radiotherapy: a clinical update. *J Intern Med.* 2016;280:203-8
- [237] Paulides MM, Verduijn GM, Van Holthe N. Status quo and directions in deep head and neck hyperthermia. *Radiat Oncol.* 2016;11:21
- [238] Vozenin MC, Bourhis J, Durante M. Towards clinical translation of FLASH radiotherapy. *Nat Rev Clin Oncol.* 2022;19:791-803
- [239] Atwell D, Elks J, Cahill K, Hearn N, Vignarajah D, Lagopoulos J, et al. A Review of Modern Radiation Therapy Dose Escalation in Locally Advanced Head and Neck Cancer. *Clin Oncol.* 2020;32:330-41
- [240] Mensour EA, Alam S, Mawani S, Bahig H, Lang P, Nichols A, et al. What is the future of treatment de-escalation for HPV-positive oropharyngeal cancer? A review of ongoing clinical trials. *Front Oncol.* 2022;12:1067321
- [241] De-Escalation Radiotherapy in Patients With Low-Risk HPV-Related Oropharyngeal Squamous Cell Carcinoma (EVADER). <https://clinicaltrials.gov/study/NCT03822897>
- [242] Mehanna H, Robinson M, Hartley A, Kong A, Foran B, Fulton-Lieuw T, et al. Radiotherapy plus cisplatin or cetuximab in low-risk human papillomavirus-positive oropharyngeal cancer (De-ESCALaTE HPV): an open-label randomised controlled phase 3 trial. *Lancet.* 2019;393:51-60
- [243] Gillison ML, Trotti AM, Harris J, Eisbruch A, Harari PM, Adelstein DJ, et al. Radiotherapy plus cetuximab or cisplatin in human papillomavirus-positive oropharyngeal cancer (NRG Oncology RTOG 1016): a randomised, multicentre, non-inferiority trial. *Lancet.* 2019;393:40-50
- [244] Machiels JP, Tao Y, Burtness B, Tahara M, Licitra L, Rischin D, et al. Pembrolizumab given concomitantly with chemoradiation and as maintenance therapy for locally advanced head and neck squamous cell carcinoma: KEYNOTE-412. *Future Oncol.* 2020;16:1235-43



Appendices

- List of publications
- PhD portfolio
- Curriculum Vitae
- Dankwoord

List of publications

- Sijtsema, ND**, Petit, SF, Poot, DHJ, Nasserinejad, K, Verduijn, GM, Warnert, EAH, Hoogeman, MS, Hernandez-Tamames, JA. Correlation between multi-delay pseudo-continuous arterial spin labeling and intravoxel incoherent motion diffusionweighted imaging in the head and neck. *Submitted*.
- Sijtsema, ND***, Lauwers, I* (*authors contributed equally), Verduijn, GM, Hoogeman, MS, Poot, DHJ, Hernandez-Tamames, JA, van der Lugt, A, Capala, ME, Petit, SF. Relating pretreatment non-Gaussian intravoxel incoherent motion diffusion-weighted imaging to HPV status and response in oropharyngeal carcinoma. *Submitted*.
- Lauwers, I, Pachler, KS, Capala, ME, **Sijtsema, ND**, Van Gent, DC, Rovituso, M, Hoogeman, MS, Verduijn, GM, Petit, SF. (2024) Ex vivo radiation sensitivity assessment for individual head and neck cancer patients using deep learning-based automated nuclei and DNA damage foci detection. *Accepted. Clinical and Translational Radiation Oncology*. <https://doi.org/10.1016/j.ctro.2024.100735>
- Verduijn, GM, **Sijtsema, ND**, van Norden, Y, Heemsbergen, WD, Mast, H, Sewnaik, A, Chin, D, Baker, S, Capala, ME, van der Lugt, A, van Meerten, E, Hoogeman, MS, Petit, SF. (2023) Accounting for fractionation and heterogeneous dose distributions in the modelling of osteoradionecrosis in oropharyngeal carcinoma treatment. *Radiotherapy Oncology* 188:109889. <https://doi.org/10.1016/j.radonc.2023.109889>
- Sijtsema, ND**, Verduijn, GM, Nasserinejad, K, van Norden, Y, Mast, H, van der Lugt, A, Hoogeman, MS, Petit, SF. (2023) Development of a local dose-response relationship for osteoradionecrosis within the mandible. *Radiotherapy Oncology* 186:109736. <https://doi.org/10.1016/j.radonc.2023.109736>
- Sijtsema, ND**, Petit, SF, Verduijn, GM, Poot, DHJ, Warnert, EAH, Hoogeman, MS, Hernandez-Tamames, JA. (2023) Multidelayer pseudocontinuous arterial spin labeling to measure blood flow in the head and neck. *NMR in Biomedicine* 36(7):e4898. <https://doi.org/10.1002/nbm.4898>
- Verduijn, GM, Petit, SF, Lauwers, I, van Norden, Y, **Sijtsema, ND**, Sewnaik, A, Mast, H, Capala, ME, Nout, RA, Baker, S, van Meerten, E, Hoogeman, MS, van der Lugt, A, Heemsbergen WD. (2023) Post radiation mucosal ulcer risk after a hypofractionated stereotactic boost and conventional

- fractionated radiotherapy for oropharyngeal carcinoma. *Acta Oncologica* 62(1):40–47. <https://doi.org/10.1080/0284186X.2022.2159772>
- Möring, MM , Mast, H, Wolvius, EB, Verduijn, GM, Petit, SF, **Sijtsema, ND**, Jonker, BP, Nout, RA, Heemsbergen, WD. (2022) Osteoradionecrosis after postoperative radiotherapy for oral cavity cancer: incidence and risk factors. *Oral Oncology* 133:106056. <https://doi.org/10.1016/j.oraloncology.2022.106056>
- Verduijn, GM, Capala, ME, **Sijtsema, ND**, Lauwers, I, Hernandez-Tamames, JA, Heemsbergen, WD, Sewnaik, A, Hardillo, JA, Mast, H, van Norden, Y, Jansen, MPH, van der Lugt, A, van Gent, DC, Hoogeman, MS, Mostert, B, Petit, SF. (2022) The COMPLETE trial: Holistic early response assessment for oropharyngeal cancer patients; Protocol for an observational study. *BMJ Open* 12(5):e059345. <https://doi.org/10.1136/bmjopen-2021-059345>
- Sijtsema, ND**, Petit, SF, Poot, DHJ, Verduijn, GM, van der Lugt, A, Hoogeman, MS, Hernandez-Tamames, JA. (2021) An optimal acquisition and post-processing pipeline for hybrid IVIM-DKI in head and neck. *Magnetic resonance in medicine* 85(2):777–789. <https://doi.org/10.1002/mrm.28461>
- de Nijs, R, **Sijtsema, ND**, Kruis, MF, Jensen, CV, Iversen, M, Perch, M, Mortensen, J. (2021) Comparison of 81mKrypton and 99mTc-Technegas for ventilation single-photon emission computed tomography in severe chronic obstructive pulmonary disease. *Nuclear medicine communications* 42(2): 160–168. <https://doi.org/10.1097/MNM.0000000000001314>

PhD portfolio

Name of PhD student:	Nienke D. Sijtsema
Erasmus MC department:	Radiotherapy Radiology and Nuclear Medicine
PhD period:	2017-2023
Promotors:	Prof. dr. Mischa S. Hoogeman Prof. dr. Juan A. Hernández Tamames
Co-promotors:	Dr. ir. Steven F. Petit Dr. ir. Dirk H.J. Poot

General courses

Basic course on regulations and organization of clinical trials (BROK)	2018
Scientific integrity	2018
Biostatistical methods I: Basic principles	2018
Data analysis with python	2019
Biomedical Scientific English Writing	2020

Specific courses

ESTRO school - Physics for modern radiotherapy	2018
Clinical radiotherapy part 1 and 2	2019-2020
"Proton therapy in the Netherlands" webinars	2021
"Deep learning in radiotherapy" webinar	2022

International conferences

ISMRM 2019	Poster	2019
ESMRMB 2019		2019
ISMRM Benelux 2020		2020
ESMRMB 2020		2020
ESTRO 2021	Poster	2021
ESTRO 2022	Oral presentation	2022

In-house seminars, workshops and presentations

Radiotherapy Journal Club including 3 presentations	2017-2022
Radiotherapy Research Days including 1 pitch, 2 presentations	2017-2022
Radiotherapy Educational Sessions	2019-2022
Radiotherapy Research Rounds	2017-2022
MR Physics seminar including multiple presentations	2017-2022

Other seminars workshop and presentations

The art of adaptive radiotherapy symposium	2018
High precision radiotherapy symposium	2019
RKF scientific project days including 1 presentation	2019-2021

Teaching

Supervising and advising bachelor student	2020
---	------

Curriculum Vitae

Nienke Sijtsema was born on May 31, 1994, in Amsterdam, the Netherlands. After receiving her high school diploma she studied Medical Natural Sciences at the VU University in Amsterdam. As part of her bachelor's degree, Nienke did a graduation internship at the department of radiotherapy at the VU University Medical Center on the accuracy of CBCT matching at zero couch rotation and X-ray verification at non-zero couch rotation in stereotactic radiotherapy. During her master's degree she further focused on medical physics and did a three-month-long internship at the Rigshospitalet in Copenhagen, Denmark, on ventilation SPECT. For her graduation internship she performed a project entitled "4-Dimensional composite spin echo for prostate 1H magnetic resonance spectroscopic imaging at 7T: A method to measure prostate cancer aggressiveness" at the imaging division of the UMC Utrecht for 9 months.



In November 2017 she started her PhD project at the Department of Radiotherapy, and the Department of Radiology and Nuclear Medicine, Erasmus MC, under the supervision of Prof. dr. Mischa Hoogeman, Prof. dr. Juan Hernández Tamames, Dr. ir. Steven Petit and Dr. ir. Dirk Poot. Her work focused on the development of response predictors for personalized radiotherapy in head and neck cancer, which resulted in this thesis.

In June 2022, Nienke started her residency in medical physics with expertise in radiotherapy at the Department of Radiotherapy, Erasmus MC.

Dankwoord

Alhoewel mijn naam op de voorkant staat, was dit boekje nooit tot stand gekomen zonder de hulp en inzet van velen. In deze sectie wil ik een aantal mensen graag in het bijzonder bedanken, te beginnen met mijn promotoren en co-promotoren.

Dr. ir. S.F. Petit, beste Steven, mede door jouw pro-actieve en positieve houding verliep het grootste deel van mijn promotietraject voorspoedig. Uiteraard hadden we veel inhoudelijke discussies, maar toch was er ook altijd oog voor het persoonlijke. Als het stroever liep, kwam je snel met out-of-the-box oplossingen (waar ik dan even aan moest wennen). Jouw enthousiasme werkt motiverend, en gaf mij de moed om door te gaan daar waar het tegen zat. Ik had me geen betere copromotor kunnen wensen, bedankt.

Prof. dr. J.A. Hernández Tamames, beste Juan, during our time together you were appointed professor. Even though you had new tasks and responsibilities to manage, you were always easy to reach and still spent many hours with me at the scanner. I think we agree that the complexity of MRI is one of the things that make it so beautiful. Thank you for sharing your knowledge of, and enthusiasm for MR physics with me.

Dr. ir. D.H.J. Poot, beste Dirk, ik waardeer jouw kritische en nauwkeurige blik enorm en ik zag mijn werk in stappen vooruit gaan op basis van jouw feedback (die vaak zo snel kwam, dat het bijna instantaan leek). Bedankt voor het delen van al jouw kennis op het gebied van MRI optimalisatie en fitting. Jouw inhoudelijke bijdrage loopt als een rode draad door deel één van deze thesis. Ik ben dankbaar dat jij als copromotor aan wilde sluiten nadat Juan aangesteld werd als professor.

Prof. dr. M.S. Hoogeman, beste Mischa, hoewel je wellicht wat minder op de voorgrond aanwezig was dan mijn andere supervisors, kon ik altijd op je rekenen als het nodig was. Ik bewonder de manier waarop je in alle rust jouw ideeën duidelijk kan overbrengen. Bedankt voor je kritische feedback, en bedankt dat je mij de kans gegeven hebt in opleiding te gaan tot klinisch fysicus.

Daarnaast wil ik de (rest van de) COMPLETE groep bedanken voor de fijne samenwerking.

Gerda, vanaf het begin was jij betrokken bij dit project. Ik heb veel van je geleerd op het gebied van hoofd-hals radiotherapie, en door jouw betrokkenheid

Dankwoord

heb ik nooit het idee gehad ver van de kliniek af te staan. Je bemoedigende woorden als er een manuscript terug kwam met een minder positief oordeel hebben mij veel goed gedaan. Bedankt voor alles.

Iris, al vrij snel nadat jij met je PhD begon werkten we veel samen. Het resultaat mag er dan ook wezen in de vorm van hoofdstuk 6 van deze thesis. Onwijs bedankt voor je enthousiasme, en je geduld als ik naast mijn nieuwe baan minder tijd had dan gehoopt. Fijn dat je als paranimf aan mijn zijde wil staan.

Marta, Martine en Ilse, bedankt voor jullie enthousiasme, inzet en gezelligheid.

A special thanks to Elekta for funding my PhD project.

Andras, bedankt voor al het werk wat je hebt verzet om onze analyses binnen Matterhorn mogelijk te maken. Ondanks de vele vragen, bleef je altijd geduldig uitleggen hoe een script precies in elkaar zat, waardoor ik vervolgens zelf verder kon.

I am grateful for my many wonderful research colleagues from the radiotherapy department. Rik, Dominique, Thyrza, Rens, Bas, Marloes, Marta, Michelle, Michelle, Laura, Sander, Marije, Nienke, Jesús, Marie, Alba, Emma, Linda, Kemal, Iva, Jannis, Wens, Sharline, Victor, Joep, Hazem and Franziska, thank you for the nice discussions, lunches and congress adventures.

I also would like to thank my MR Physics colleagues Laura, Alexandra, Dori, Theresa, Piotr, Gyula, Samy, Riwanj, Chaoping, Willem, Emanoel, Krishnapriya, Anita, Esther, Joost, Yulun and Fatemeh. Thank you for your valuable input during seminars, and your welcoming demeanor, even though I wasn't around much. A special thanks to Krishnapriya for spending so many hours with me at the scanner.

Joan en Maarten, ondanks dat mijn promotie nog niet afgerond was, gaven jullie mij de kans om onder jullie begeleiding in opleiding te gaan tot klinisch fysicus. Bedankt voor jullie vertrouwen, interesse, geduld en flexibiliteit.

Jacqueline en Jolanda, bedankt voor al jullie hulp bij kleine en grote praktische zaken.

Beste KFMers en versnellertechnici, in de tijd dat ik mijn proefschrift indiende, liep ik bij jullie stage en hebben jullie mij meer dan eens diep horen zuchten. Bedankt voor jullie interesse en bemoedigende woorden.

Lieve vrienden en familie, bedankt voor jullie interesse door de jaren heen. Het uitleggen van mijn werk is een uitdaging en ik waardeer jullie inspanningen om het toch te proberen te begrijpen.

Kim, bedankt voor je luisterend oor, adviezen en het delen van ervaringen door de jaren heen. Het is fijn om mijn ups en downs te kunnen delen met een gelijkgezinde. Ik ben blij dat jij als paranimf aan mijn zijde staat.

Femke, bedankt voor het maken van de omslagillustratie, waardoor mijn proefschrift een heel persoonlijke tint heeft gekregen. Alhoewel wij elkaar niet vaak zien, weet ik dat het goed zit.

Lieve mama, je hebt al mijn successen en tegenslagen voorbij horen komen. Bedankt voor je geduld en verhelderende perspectieven, en je begrip als er weinig tijd was om langs te komen. Ik waardeer het enorm dat je altijd voor mij klaarstaat, zelfs voor de meest onbenullige dingen.

Lieve Patrick, jouw rust bracht mij ook rust. Bedankt dat je, ook zonder een vraag van mijn kant, me veel uit handen nam zodat ik dit avontuur tot een goed einde kon brengen. Ik kijk uit naar meer vrije tijd samen!

Papa, Ernst, jullie waren te kort in mijn leven, toch draag ik jullie levenslessen bij me. Aan jullie draag ik dit proefschrift op.

