MR image biomarkers of the parotid gland stem cell rich region to predict xerostomia after radiotherapy

Suzanne de Vette

UNIVERSITY OF TWENTE.

umcg

MR image biomarkers of the parotid gland stem cell rich region to predict xerostomia after radiotherapy

Suzanne de Vette

A thesis submitted for the degree Master of Science

December 20, 2021

Technical Medicine
Medical Imaging and Interventions
Faculty Science and Technology
University of Twente
Enschede, the Netherlands

Graduation committee

Technical supervisor (chairman)
Prof. dr. Ir. C.H. Slump

Department of Robotics and Mechatronics
University of Twente
Enschede, The Netherlands

Medical supervisor
Dr. R.J.H.M. Steenbakkers

Department of Radiotherapy
University Medical Center Groningen
Groningen, The Netherlands

Daily supervisor
M.I. van Rijn-Dekker MSc.

Department of Radiotherapy
University Medical Center Groningen
Groningen, The Netherlands

Technical supervisor
Dr. Ir. N.M. Sijtsema

Department of Radiotherapy
University Medical Center Groningen
Groningen, The Netherlands

Process supervisor
R.M. Krol, MSc.

TechMed Center
University of Twente
Enschede, The Netherlands

External member
Prof. dr. Ir. B. ten Haken

Department of Magnetic Detection & Imaging
University of Twente
Enschede, The Netherlands

Preface

Before you lies my master's thesis for the study Technical Medicine, track Medical Imaging and Interventions performed at the University of Twente. My thesis concerns the addition of MRI image biomarkers of the stem cell rich region of the parotid gland to a clinical reference model for the prediction of daytime xerostomia. I conducted this research at the Radiotherapy department of the University Medical Center Groningen.

I would like to thank everybody who helped me achieve writing this thesis. First, I would like to thank my daily supervisor, Irene van Rijn-Dekker, for her guidance and helping hand in letting me discover predictive modelling. Furthermore, I would like to thank Roel Steenbakkers for helping me grow my self-confidence in the clinic, which has resulted in me to find great joy in the clinical part of my internship. Further, I would like to thank Marianna Sijtsema for guiding me in the technical part of my internship and her encouragement to make my own choices. Additionally, I would like to thank Kees Slump for his critical eye for detail and his broad guidance during my internship. Furthermore, I would like to thank Ruby Krol for encouraging me to grow as a person. At last, I would like to thank Sanne van Dijk for her help with retaining realistic goals and giving me more insight in what could lie in the future as a technical physician.

I hope you enjoy reading!

Suzanne de Vette

Groningen, December 20, 2021

Abstract

Purpose - To evaluate the addition of Magnetic Resonance Imaging (MRI) image biomarkers (IBMs) of the parotid gland stem cell rich (SCR) region to a clinical reference model to predict daytime xerostomia 12 months after radiotherapy for head and neck cancer (HNC).

Methods and Materials - A retrospective analysis was performed on 104 HNC patients who were treated using curative radiotherapy between 2018 and 2020. T1 turbo spin echo MRI scans, planning Computed Tomography (CT) scans, dose distributions and GRIX daytime xerostomia grade (baseline and 12 months follow-up) were prospectively collected. SCR regions were delineated on CT. Parotid gland and SCR region structures were transferred to MRI scans, MRI scans were normalized (standardization by intensities), and IBMs were extracted from the ipsilateral and contralateral SCR region. Dose parameters were determined based on the dose distribution. Pre-selection of IBMs was executed using the Bayesian Information Criterion and Spearman correlation with the endpoint (<0.8). Logistic regression sub-models were created based on predictor groups with intercorrelation of <0.8 and sub-models were combined in a composite model. Internal validation was executed by bootstrapping 100 times. Selected IBMs were added to a clinical reference model and using a likelihood-ratio test, the addition of IBMs to the model was tested.

Results - Predictive IBMs were the long run low grey level emphasis ipsilateral and the short run high grey level emphasis contralateral. The area under the curve (AUC) for this model was 0.68 (0.43-0.86). After internal validation, the AUC decreased to 0.57 (0.45-0.68). The clinical reference model had an AUC of 0.55 on our study population after closed-testing procedure and the likelihood-ratio test revealed that adding the IBMs did not improve this model.

Discussion - With another MRI sequence specifically created for sialography and subsequent a segmentation of the SCR region based on the main ducts of the parotid gland, the SCR region could be defined more accurately, most likely resulting in more predictive IBMs. Furthermore, due to differences in the study population, the clinical reference model did not have a good fit on our study population. These differences were probably caused by changes in treatment planning and technological progress over the years. In conclusion, it was found that MRI IBMs of the SCR region can be predictive of daytime xerostomia 12 months after radiotherapy, but more research is needed.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Explanation	Abbreviation	Explanation
AUC	Area Under the Curve	IBM	Image Biomarker
BIC	Bayesian Information Criterion	IMPT	Intensity-Modulated Proton Therapy
BSL	Baseline	IQR	Inter Quartile Range
CI	Confidence Interval	LRLGE	Long Run Low Grey Level Emphasis
CT	Computed Tomography	MRI	Magnetic Resonance Imaging
D40	Dose that 40% of the volume receives	MICE	Multivariate Imputation by Chained Equations
D_{mean}	Mean dose the volume receives	NGTDM	Neighborhood Grey-Tone Difference Matrix
DNA	Deoxyribonucleic acid	NTCP	Normal Tissue Complication Probability
DVH	Dose-Volume Histogram	OR	Odds Ratio
EBV	Eppstein-Barr virus	P90	90^{th} intensity percentile
EDAR	Ectodysplasin A receptor	PET	Positron-Emission Tomography
FDA	US Food and Drug Administration	PG	Parotid gland
GLCM	Grey Level Co-occurence Matrix	QUANTEC	Quantitative Analyses of Normal Tissue Effects in the Clinic
GLRLM	Grey Level Run-Length Matrix	ROS	Reactive Oxygen Species
GLZSM	Grey Level Size Zone Matrix	SCR	Stem Cell Rich
GRIX	Groningen Radiotherapy-Induced Xerostomia	SG	Submandibular gland
Gy	Gray	SRE	Short Run Emphasis
HL	Hosmer-Lemeshow	SRHGE	Short Run High Grey Level Emphasis
HNC	Head- and Neck Cancer	TNM	Tumour Node Metastasis
HPV	Human-Pappiloma virus	UMCG	University Medical Centre Groningen
IBSI	Image Biomarker Standardization Initiative	VMAT	Volumetric Modulated Arc Therapy

Chapter 1
Background information

1.1 Head- and neck cancer

According to *Fitzmaurice et al. (2019)*, the estimated worldwide incidence of head- and neck cancer (HNC) was 890,000, representing 5.3% of all cancers [1]. In their and our study, HNC is comprised of cancer in the oral cavity, nasopharynx, oropharynx, hypopharynx and larynx (see Figure B.1). The mortality rate of HNC was 5.3% of all cancer-related deaths world-wide.[1] The incidence of HNC in the Netherlands was the highest in the age group of 65-74 years, with approximately one-third of the incidence in the past 5 years in the Netherlands [2]. However, in recent years, an increase in HNC as a result of the Eppstein-Barr virus (EBV) (nasopharyngeal area) and Human-Papilloma virus (HPV) (oropharyngeal area) was found, which has targeted mostly younger patients in the age group of 60-64 years.[3, 4]

The main risk factors for the development of HNC include tobacco exposure and alcohol use. [6, 7] Furthermore, presence of the EBV or HPV and gastroesophageal reflux can cause changes to the mucosal cells, which is also a risk factor for developing HNC. [8, 9]

1.2 Radiotherapy

HNC patients are treated with surgery, (concurrent chemo)radiotherapy or a combination of these.[10] Radiotherapy is applied when the tumour is inoperable or radical resection margins of more than 5 mm cannot be guaranteed without loss of function, when there are positive surgical margins, or when radiotherapy is beneficial for the preservation of function of organs.[11] Subsequently, radiotherapy is the

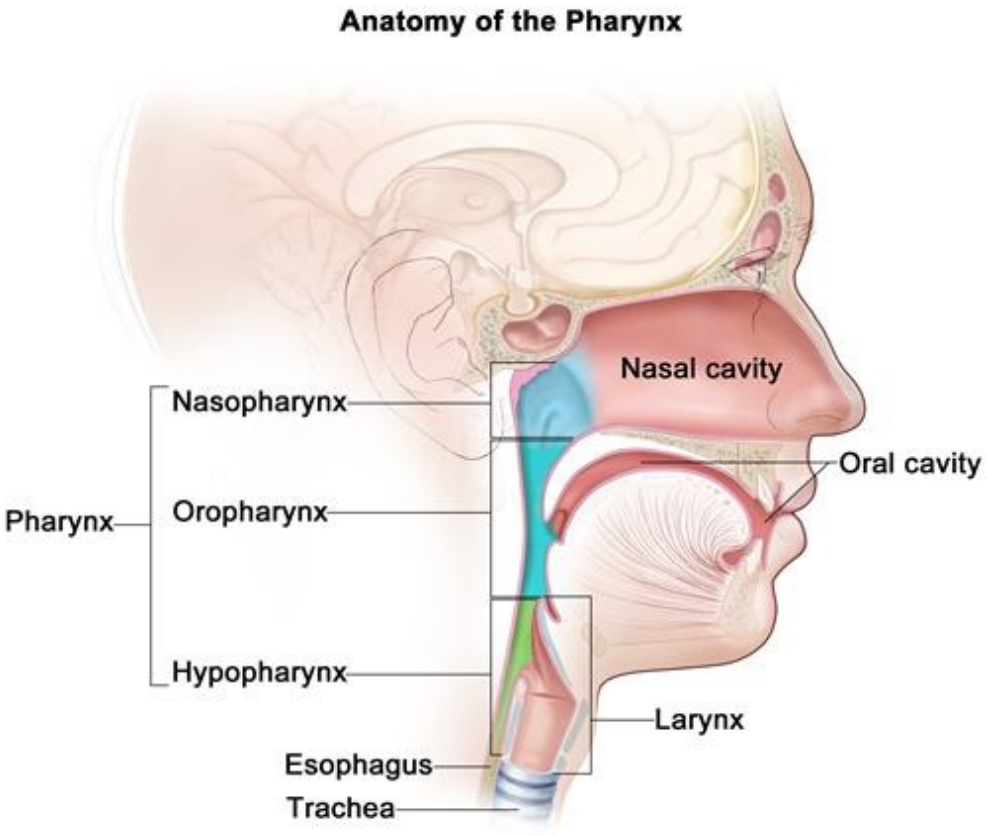


Figure B.1: Overview of head and neck cancer regions. The head and neck area can be subdivided into: oral cavity, nasopharynx, oropharynx, hypopharynx and larynx. [5]

treatment option of choice in 70% of the patients with HNC. The goal of radiotherapy is to kill cancer cells and thereby removing the cancerous tissue to cure patients.

Even though radiotherapy can preserve functional tissue better than surgery, radiotherapy can cause several side effects. The sensation of dryness of the mouth, called xerostomia, is the most often reported side effect of radiotherapy and has a large impact on the quality of life [12–14]. Xerostomia is caused by radiation damage to the salivary glands, causing salivary gland dysfunction. Other side effects include i.e. difficulty swallowing (dysphagia), dryness, redness and blistering of the skin and mucosa, soreness of the throat and mouth, hoarseness, loss of taste, loss of smell and sticky saliva. Many factors contribute to the development of side effects, such as the location of the tumour, the type of radiation (photon/proton), the daily fraction size, whether irradiation of the elective lymph node levels is necessary, whether the patient already had complaints due to the tumour itself and more.[11]

At the radiotherapy department of the University Medical Centre Groningen (UMCG) in the Netherlands, several protocols are used to treat patients depending on patient and tumour characteristics, according to the national guidelines. In general, patients who receive primary radiotherapy will receive 70 Gray (Gy) to the tumour and 54.25 Gy to the elective lymph node areas (if necessary) in 35 fractions, which are administered on five days a week. [15] However, the method of treatment also depends on the TNM stage of the tumour and the age of the patient. For tumours which have a tumour staging of T1 or T2 and node staging of N0, accelerated radiotherapy is applied when the patient is younger than 70 years of age. This means that the patient receives 6 fractions of 2 Gy administered on 5 days a week (so one day with 2 fractions), with a total dose of 70 Gy. When no contra-indications are present, concomitant radiotherapy with cisplatin/carboplatin/5 FU or accelerated radiotherapy in combination with cetuximab are treatment options for patients of 70 years or younger who have another tumour or node staging.[16–25]

Radiotherapy aims to deliver 70 Gy to the tumour area, 54.25 Gy to the elective lymph node area and as low dose as possible to other structures to prevent side effects and complications. At the radiotherapy department of the UMCG, it is possible to irradiate HNC patients with protons or photons. Due to the difference in physical properties between protons and photons, the peak dose per beam can be set on a certain depth in proton therapy, whereas the peak dose of photon therapy lies at the surface of the body, which causes proton therapy to be able to have a steep dose fall-off and thus a lower dose to healthy tissue. To determine which patients benefit from proton therapy, Normal Tissue Complication Probability (NTCP) models are created to predict the toxicity of both the photon and proton irradiation plan. [26, 27] A patient qualifies for proton therapy when the amount of xerostomia or dysphagia can be reduced by a certain percentage (5% for grade 3 complications and 10% for grade 2 complications) when proton therapy is applied instead of photon therapy. [28]

1.3 Parotid gland and stem cell rich area

During radiotherapy, the parotid gland tissue undergoes multiple changes. A volume loss of up to 35% during the course of the therapy is reported [29–36]. In a review by *Jasmer et al. (2020)*, it is described how salivary gland tissue is damaged by irradiation [37]. The factors that contribute to acute dysfunction of the salivary glands most are aberrant calcium signaling, DNA damage, reactive oxygen species (ROS) formation and rapid apoptosis of acinar cells (both mucous and serous cells). Factors that furthermore contribute to chronic salivary gland dysfunction are inflammation, changes in neuronal and vascular tissue, senescence of the progenitor cells, a change in the cytoskeleton and fibrotising of the salivary gland. [37] *Braam et al. (2005)* showed that the parotid gland tissue regenerates over time, since the salivary flow rate after 5 years had increased 32% in comparison with the salivary flow rate 1 year after radiotherapy [38].

Saliva production is conducted by the mucous and serous cells in the salivary glands. Mucous and serous cells have a slow turnover rate, suggesting that they are less radiosensitive in comparison to cells with a high turnover rate. However, the amount of side effects suggests that the parotid gland is more radiosensitive than expected based on the radiosensitivity of these cell types.[39] Therefore, a study by *Pringle et al. (2013)* suggested that the stem cells and progenitor cells in the salivary glands get sterilized by the radiation, causing them to be unable to maintain their self-renewal and differentiation properties [40]. The stem cells and progenitor cells of the parotid gland are concentrated along the main ducts of the parotid gland, as can be seen in Figure B.2, further called the stem cell rich (SCR) region [41]. This is in

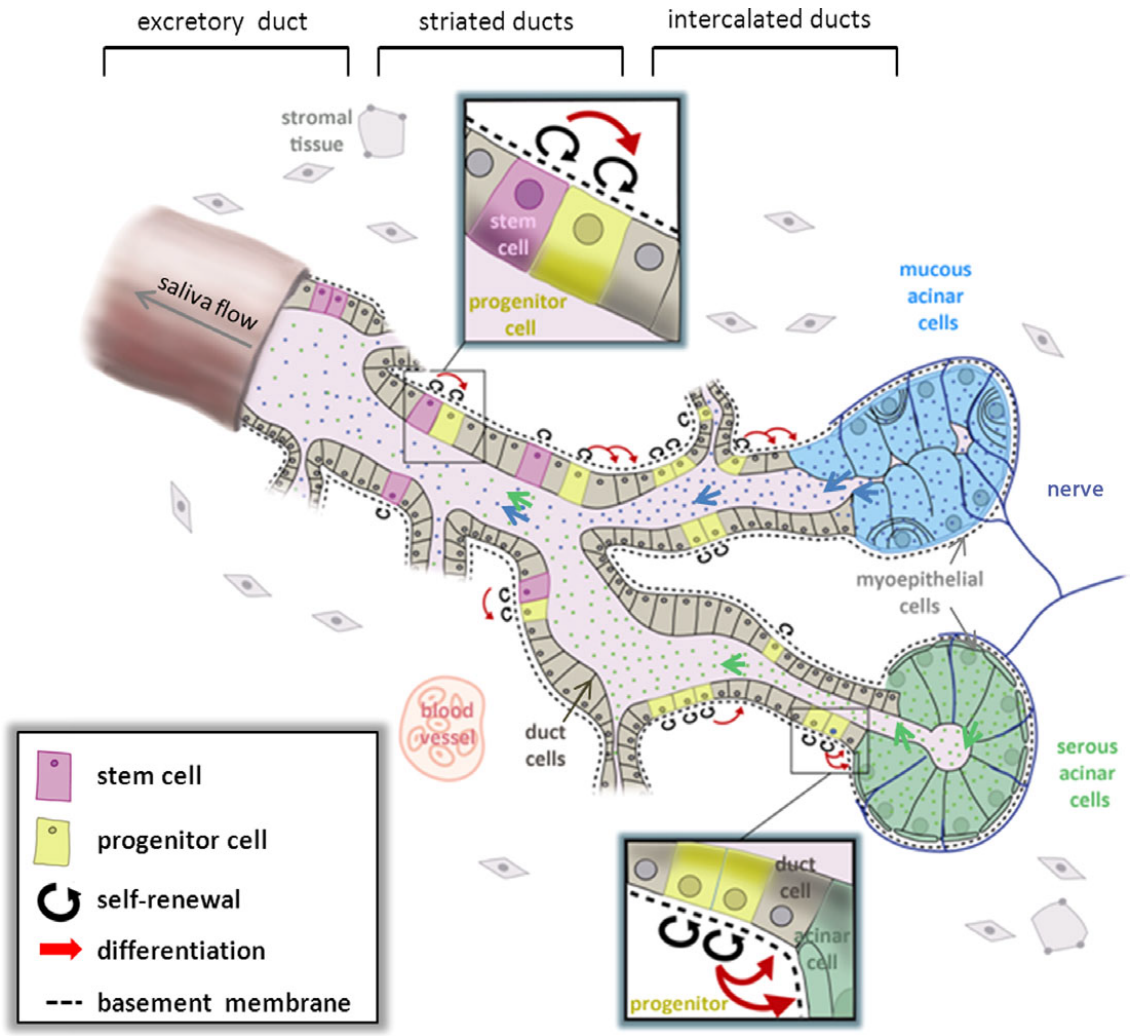


Figure B.2: Parotid gland structure on cellular level.[40]

.

line with a study by *Konings et al. (2005)* towards the region-dependent radiosensitivity of the parotid gland, in which it is concluded that the caudal part of the parotid gland is less radiosensitive, where no main ducts of the parotid gland are present. In this study, it was found that regeneration of parotid tissue is highly dependent on the damage caused to the cranial part (lateral and part of dorsal lobe) of the parotid gland. [42] This suggests that sparing the cranial part of the parotid gland where stem cells and progenitor cells reside may have a large impact on the extent of xerostomia complaints.

Currently, multiple methods are applied to reduce the damage to salivary glands, such as salivary gland sparing radiotherapy [43]. In the study by *Eisbruch et al. (2001)*, it was found that sparing the parotid and submandibular glands resulted in a long-term clinical benefit for the patients, indicating that the impairment of salivary production was decreased. According to *Hawkins et al. (2017)*, patients treated with this technique indeed produce more saliva than patients treated with standard radiotherapy, however, the patient-rated xerostomia only improved marginally [44]. In the study by *Hawkins et al. (2017)*, it was furthermore found that sparing not only the parotid gland, but all salivary glands did improve patient-rated xerostomia, which is in line with other studies [13, 14, 29, 44]. Furthermore, *Beetz et al. (2014)* tested the QUANTEC group guidelines which are created to prevent moderate-to-severe xerostomia. When the criteria of the QUANTEC guidelines can be met, the risk of developing moderate-to-severe xerostomia reduced with 20% in younger patients without pre-existing xerostomia before the treatment.[45] All of these studies did find a positive effect on reducing xerostomia, but often this was in specific subgroups

(high Karnofsky score, low age groups, no xerostomia present before radiotherapy), which the majority of HNC patients do not belong to. Thus, a better way of reducing the risk of xerostomia in all patient groups has yet to be found.

Other methods to protect the parotid glands from radiation damage is by administering medication to protect tissue from radiation damage before the treatment starts. Currently, the only FDA approved radioprotective therapeutic is amifostine. Amifostine is believed to be a ROS scavenger in non-tumorous tissue. ROS are mainly responsible for DNA damage, causing apoptosis. With amifostine, less DNA damage is achieved in healthy tissue, leading to less side effects of the therapy. [46, 47] However, more recent studies did not find a benefit in using amifostine over the use of a placebo [48, 49]. Some preclinical studies to other therapeutics show promising results in protection against radiation damage for the therapeutics A-438079, dasatinib and imatinib [50, 51]. However, these therapeutics have currently not proven to be effective in clinical studies. To further reduce xerostomia, pharmacological options such as rapamycin and EDAR agonist monoclonal antibodies are considered for regeneration of the salivary gland [52–54].

Another promising method to reduce xerostomia is stem cell transplantation of the parotid gland stem cells [40]. To our knowledge, the change in human SCR regions due to delivered dose has not been investigated yet. The X-Prevent project addresses the change in parotid SCR region in relation to xerostomia to give insight in how xerostomia can be prevented. Eventually, it will be evaluated which patients can be helped with a stem cell transplantation.[55] The effect of dose to the SCR region in relation to xerostomia is investigated in a double-blind randomized controlled trial, in which it became evident that dose to the SCR region was more predictive for xerostomia than dose to the whole parotid gland.[56] This Master project is a part of the X-Prevent project.

1.4 Image Biomarkers

The parotid gland tissue undergoes changes during and after irradiation. These changes can be detected by image biomarkers (IBMs) and some of them can be related to xerostomia [57–67]. The specific change of IBMs in the SCR regions has not yet been an area of interest, despite the importance of these areas in reducing xerostomia after radiotherapy [40, 41, 56]. Hypothetically, changes in this area could be even better predictors for xerostomia than changes in the whole parotid gland.

The first mention of IBMs in baseline imaging of the parotid gland in the prediction of xerostomia is in a study by *Nakatsugawa et al. (2016)* [63]. In this study, several IBMs are shown to be a valuable addition to dose volume histogram (DVH) data of the combined parotid and submandibular glands in predictive modeling of radiation-induced xerostomia. These IBMs were the texture parameters of the ipsilateral parotid gland and the texture and shape parameters of the ipsilateral submandibular gland. The area under the curve (AUC) of only DVH data was 0.7, while the AUC of the DVH data in combination with the IBMs was 0.85.

Hereafter, studies by *van Dijk et al. (2016-2018)* identified more IBMs related to the development of radiation-induced xerostomia. They showed that parotid gland surface reduction on CT and the baseline xerostomia score are correlated with moderate-to-severe xerostomia 6 to 12 months after radiotherapy [57]. Furthermore, a model exists which predicts the parotid shrinkage after radiotherapy. However, this model is not linked to xerostomia.[66]

Since it is of more clinical value to determine the chance of developing xerostomia before the treatment starts, a second study was conducted using IBMs of the parotid gland in pre-treatment CT scans. This study demonstrated the IBM which significantly improved the baseline model (based on DVH parameters and baseline xerostomia) is the 'Short Run Emphasis' (SRE) of the 'grey level run-length matrix' (GLRLM), a textural IBM. An AUC of 0.76 was achieved, which is lower than the study by Nakatsugawa et al. (2016)[63]. However, the study by *van Dijk et al. (2017)* included almost three times more patients (87 vs 249), which makes these results more convincing [58]. A high SRE value indicates heterogeneous parotid tissue, which implies that there is a fatty tissue deposition in the parotid glands. In Sjögren's syndrome, in which patients also have fatty tissue in the parotid gland, it was shown that the amount of fatty tissue in the parotid gland has a relationship with parotid function impairment [68]. To further investigate the heterogeneity of the parotid tissue and its influence on the development of xerostomia, a subsequent study was executed based on ^{18}F-FDG PET-CT IBMs of the parotid glands. The 90th

percentile of standardised uptake values, an intensity IBM, and the 'Long Run High Grey-Level Emphasis 3', a textural IBM derived from the GLRLM, were the most predictive for xerostomia, with an AUC of 0.76 and 0.75 respectively. These IBMs were found to be correlated and when combined, they do not add independent information. The found IBMs can both be related to the amount of fatty, non-functional parotid tissue, supporting the findings by *van Dijk et al. (2017)*. [58, 60] To investigate the influence of fatty tissue in the parotid gland even further, MRI IBMs are analyzed. The IBMs that were most predictive of xerostomia are the 90th percentile of the normalized MRI intensity in the parotid gland and the 'grey level non-uniformity normalized', derived from the GLRLM. Again, these IBMs are correlated and do not add individual information to the model. The AUC for addition of both IBMs independently to the baseline model was 0.83 for both IBM-models. [59]

During the study of *van Dijk et al. (2017)*, another research group also found IBMs correlated with xerostomia [64]. This study used several highly correlated parameters in their model, which made the AUC relatively high (0.91). However, including highly correlated parameters may cause a large standard error of their model, creating an unstable model [69]. Furthermore, only k-fold validation (with k=4) is applied, while *van Dijk et al. (2017)* applied external validation, which is more reliable. [58, 64] Both *van Dijk et al. (2018)*, *Wu et al. (2018)* and *Rosen et al. (2018)* compared IBMs of CT scans created during treatment [61, 65, 67]. The most important factors in the prediction of xerostomia were the change in parotid gland surface area between the start of the treatment and week 3 of the treatment (AUC = 0.93) [61] and changes in mean intensity on CT of the parotid glands and parotid volume between the start of the treatment and week 3 or 5 of the treatment (Pearsons correlation coefficient = 0.71, success rate = 93% [67] and AUC = 0.78 [65]. These studies did have different endpoints (moderate-to-severe xerostomia 12 months after treatment [61, 65] and severe xerostomia directly after treatment [67]), but their results seem consistent with each other.

To this day, only 2 studies have focused on IBMs of parotid glands on pre-treatment T1 turbo spin echo MRI scans and their relation with the possibility to develop xerostomia [59, 62]. Differences in these works include the usage of IBMs of MRI scans before [59] or after gadolineum administration [62], wavelet filtering of the scans [62], normalization of the MRI scans [59] and the usage of a standard model to which the MRI IBMs are added to predict xerostomia [59]. In the article by *Sheikh et al. (2018)*, multiple models are created which are based on a mix of DVH data, clinical data, CT IBM data and MR IBM data [62]. Because of this, the models are not fully comparable. The best performing model was the model containing all data (AUC = 0.79), while the best performing model in the study by *van Dijk et al. (2018)* had an even higher AUC of 0.83 for their model [59]. Surprisingly, all MR IBMs that are selected by *Sheikh et al. (2018)* resulted from the wavelet filtered MR images, a method of image manipulation that was not applied by *van Dijk (2018)*. Due to the inter-scan variability that is introduced by the MR modality, normalization is applied by *van Dijk et al. (2018)*, while this has not been executed by *Sheikh et al. (2018)*. This could have a negative effect on the model by *Sheikh et al. (2018)*.

1.5 Aim and key objectives

As stated, many different approaches to use imaging as a predictor of radiation-induced xerostomia have been applied, but none of them specifically used the SCR region of the parotid gland as region of interest. The next step is to investigate whether IBMs of the SCR region of the parotid glands on MRI have a comparable or even better predictive value in the prediction of xerostomia. The purpose of this study is to investigate the addition of MRI IBMs of the SCR region of the parotid gland to a clinical linear regression model to predict xerostomia after radiotherapy.

Several steps have to be undertaken to achieve this goal;

1. Data collection
   (a) Inclusion of patients who underwent radiotherapy for HNC, who received an MRI before treatment.
   (b) Check on each sequence whether the parotid gland is visible.
   (c) Select MRI sequence for study based on properties of the sequence and availability.

   (d) Transfer the delineated structure for the parotid gland and the SCR region from planning CT to MRI.

2. Normalize MRI, standardization of intensities
3. Extract intensity and textural radiomics
4. Imputation for missing clinical data
5. Model creation
   (a) Model based on radiomics of the SCR regions.
   (b) Addition of selected radiomics to clinical model.
6. Model outcome interpretation and comparison

Chapter 2
MR image biomarkers of the parotid gland stem cell rich region to predict xerostomia after radiotherapy

2.1 Introduction

Xerostomia is the most common side-effect of radiotherapy for head and neck cancer (HNC), with furthermore a large impact on daily life [44, 70–72]. Xerostomia is caused by radiation damage to the salivary glands, leading to salivary gland dysfunction [37]. To gain more insight in the predicting factors of xerostomia, normal tissue complication probability (NTCP) prediction models are created. Various potential predictive parameters were investigated; clinical parameters, answers to questionnaires, dosimetric parameters of the treatment plan and image biomarkers (IBMs) of several image modalities. [26, 43, 57–67, 73–77] IBMs are mathematical characteristics of a region of interest in patient's scans, such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI) or Positron-Emission Tomography (PET) scans [78].

Studies that investigated the role of IBMs in the prediction of xerostomia were conducted based on several image modalities, at different time points of the treatment period [57–66, 73]. Due to its superior soft-tissue contrast, MRI scans are expected to provide more information about structural parameters within the parotid gland than CT or PET imaging [79]. This is supported by a comparison of results by *van Dijk et al. (2017-2018)*, which showed that IBMs based on MRI scans give a better predictive value (area under the curve (AUC) = 0.83) than IBMs based on CT (AUC = 0.76) or PET imaging (AUC = 0.76)[58–60]. Currently, only two studies have been performed that look into MRI IBMs in the prediction of xerostomia [59, 62]. Scans in both studies have been created with the T1 turbo-spin echo sequence. However, differences in methods of modelling, endpoints, usage of the gadolineum contrast-agent and method of post-processing the MRI scans make direct comparison difficult. The most predictive parameters for the best performing models of *Sheikh et al. (2019)* (model based on clinical parameters, dosimetric parameters, CT IBMs and MRI IBMs) and *van Dijk et al. (2018)* (model based on clinical parameters, dosimetric parameters and MRI IBMs) can be found in Table 1 [59, 62]. The model with the largest AUC after external validation was the model with the textural IBM by *van Dijk et al. (2018)*, with an AUC and confidence interval (CI) of 0.83 (0.67-0.99).

At present, there is still a large uncertainty in the current prediction models for xerostomia that are based on clinical factors, dosimetric parameters and IBMs of the salivary glands to such an extent that these models cannot be used clinically. This can be partially due to the radiosensitivity of the parotid gland, which is higher then expected from glandular tissue [39]. In a later study by *Konings et al.*, a relation was found between different areas of the parotid glands and their radiosensitivity [42]. Hereafter, the areas with a higher radiosensitivity were correlated to the areas where stem cells reside, that is to say in the region of the parotid gland containing the major ducts [41]. Sterilization of the stem cells results in a decreased ability to recover the complete parotid gland after radiotherapy, thus causes impaired parotid function [40]. In clinical practice, the mean dose to the stem cell rich (SCR) region was found to be the best predictor for patient-rated daytime xerostomia and physician-rated grade $\geq$2 xerostomia. [56]

At present, no studies have been published that investigate the effect of IBMs of the SCR region of the parotid gland in predictive modeling of radiotherapy-induced xerostomia. From a radiobiological perspective, the SCR region may give more insight in the development of xerostomia. Therefore, this study will investigate whether addition of MRI IBMs of the SCR region to a model created to predict xerostomia will improve the prediction of xerostomia.

2.2 Methods and Materials

2.2.1 Study population

The patient population consisted of patients who received primary radiotherapy of 70 Gy in the head-and-neck area, received in 6-7 weeks at the University Medical Centre Groningen (UMCG). Furthermore, these patients must have received an MRI scan for treatment planning between January 2018 and April 2020. Moreover, parotid glands must be visible on the MRI scans and the SCR region must be detectable. All patients were treated with Intensity-Modulated Proton Therapy (IMPT) or Volumetric Modulated Arc Therapy (VMAT). Treatment plans were generated by the protocol that is used for HNC patients in the Radiotherapy department of the UMCG. This included reducing radiation dose to the parotid glands, submandibular glands, oral cavity and pharyngeal constrictor muscles as much as possible without compromising the dose to the target volume.

Table 1: Comparison predictive models xerostomia with MRI IBMs.[59, 62]

Sheikh et al. (2019)[§]			*van Dijk et al. (2018)*		
Parameter class	Parameter	OR/p-value	Parameter class	Parameter	OR/p-value
Dose	D40 contralateral PG	2.79 / 0.04	Dose	D_{mean} PG	-*/-*
CT texture IBM	ipsilateral SG GLSZM Grey Level Non-uniformity Normalized	2.29 / 0.04	Clinical	Baseline moderate-to-severe xerostomia	-*/-*
MRI texture IBM	ipsilateral SG GLSZM Small Area High Grey Level Emphasis	3.59 / 0.002	MRI intensity IBM	PG P90[‡]	1.03 / 0.004
MRI texture IBM	ipsilateral SG GLSZM Grey Level Non-uniformity Normalized	0.40 / 0.04	MRI texture IBM	PG GLRLM Grey Level Non-uniformity Normalized[‡]	0.34 / 0.004
AUC (CI)[†] = 0.79 (0.78-0.80)			AUC (CI)[†] P90 = 0.88 (0.79-0.96) AUC (CI)[†] GLRLM = 0.88 (0.79-0.96)		

Abbreviations: OR = Odds Ratio, D40 = dose that 40% of the volume receives, D_{mean} = mean dose the volume receives, PG = Parotid Gland, SG = Submandibular gland, GLSZM = Grey Level Size Zone Matrix, P90 = 90^{th} intensity percentile of normalised MRI-units, GLRLM = Grey Level Run Length Matrix, CI = Confidence Interval.

[§] In the article by *Sheikh et al. (2019)*, all selected IBMs were obtained from wavelet filtered images.

* These parameters come from a reference model [26, 27], in which the AUC was 0.81 (95% CI: 0.71–0.91), and the R^2 was 0.39 on this dataset.

[†] Values are for unvalidated models, since validation methods differ.

[‡] These parameters were both added separately to the base model. Together, they did not add more information since they were highly correlated with each other.

Exclusion criteria were for patients to be having a salivary gland tumour or a skin tumour. Furthermore, patients who had less than 1 year follow-up at the time of the data retrieval were excluded. Moreover, patients who did not fill in any questionnaires regarding toxicity were excluded. Additionally, patients were excluded if they underwent surgery in the head-and neck region or radiotherapy prior to treatment. MRI scans were visually checked for noise or artefacts that would hinder determination of IBMs.

2.2.2 MRI acquisition and normalization

All MRI scans were acquired before treatment for tumour delineation purposes. T1-weighted MRI scans were made using a Siemens MAGNETOM Prisma 3T scanner (Siemens Healthineers, Erlangen, Germany) with a turbo spin echo sequence (TE = 15 ms, TR = 875 ms, flip angle = 70 degrees, matrix size = 320 x 320, pixel size = 0.3 x 0.3 mm, slice thickness = 3 mm). No intravenous contrast agents are given to the patient prior to scanning and no fat suppression has been conducted during scanning.

The intensity values on MRI scans are not comparable between scans due to inhomogeneity in the magnetic field, placement of the radiofrequency receiver coils and other factors.[80] Therefore, it is advised to normalize MRI scans prior to the IBM extraction. In order to achieve this, the properties of the T1 turbo spin echo sequence were used. On this sequence, fatty tissue has one of the highest intensity values and can be located easily. Therefore, fatty tissue of the cheek is delineated in 5 slices of each MRI scan. Hereafter, the standardization of intensities technique proposed by *Robitaille et al. (2012)* is applied. [81] In this technique, the image values were multiplied by a constant factor (here arbitrarily chosen to be 350) and divided by the average cheek fat tissue intensity value. An example of the normalization process can be found in Figure 1. The resulting grey scale was from 0 to 350.

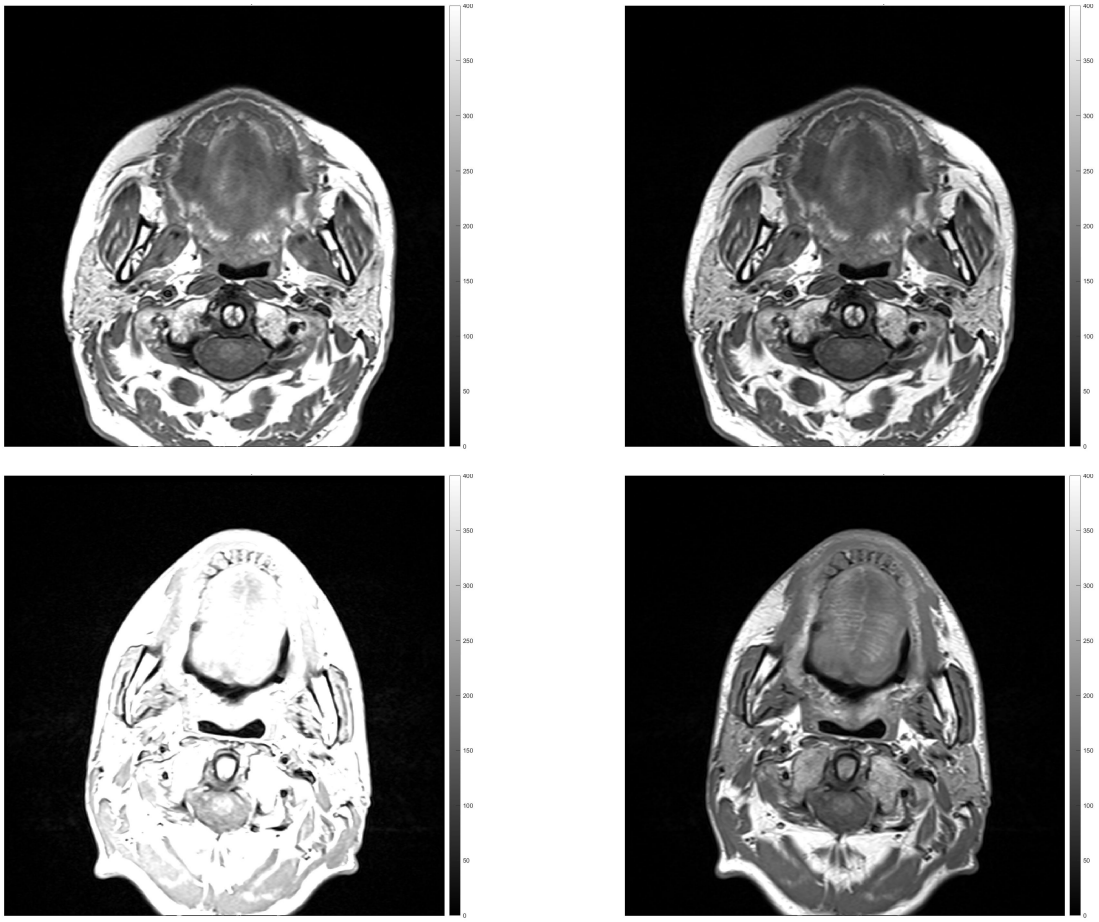
2.2.3 Definition structures

Prior to treatment, the parotid glands were delineated on CT scans based on the method by *Brouwer et al. (2015)*.[82] Based on these contours, the SCR region is defined as the area around the main salivary ducts of the parotid gland [41, 56]. This area is defined in the following manner: firstly, the transversal slice on which the parotid gland's centre is found based on the contour data using MATLAB (Mathworks, 2018a, the MathWorks, Inc., Natick, Massachusetts, 2018). Hereafter, the intersection of the parotid gland with the mandible and the masseter muscle on this transversal slice is found in the Raystation Klinisch 10B treatment planning system (RaySearch Laboratories AB, Stockholm, Sweden). Subsequently, this point of interest is shifted 0.5 cm in posterior-lateral direction. Hereafter, a volume around this point of interest is created by expanding it with 1 cm in the anterior-posterior direction and 2 cm in the cranial-caudal direction. At last, the part of the generated volume within the parotid gland is defined as the SCR region. An example of the SCR region can be found in Figure 2.

The parotid gland structure and the SCR region structure were rigidly transferred from planning CT scan to the planning MRI scan with software of Mirada (Mirada Medical Ltd., Oxford, United Kingdom). These delineations on the MRI scans were manually checked and adjusted if necessary. When there were large deviations between the contour on the CT scan and the visible structure on the MRI scan, the SCR region was redefined based on the data extracted from MATLAB used to define the SCR region in the first place.

2.2.4 IBM candidates

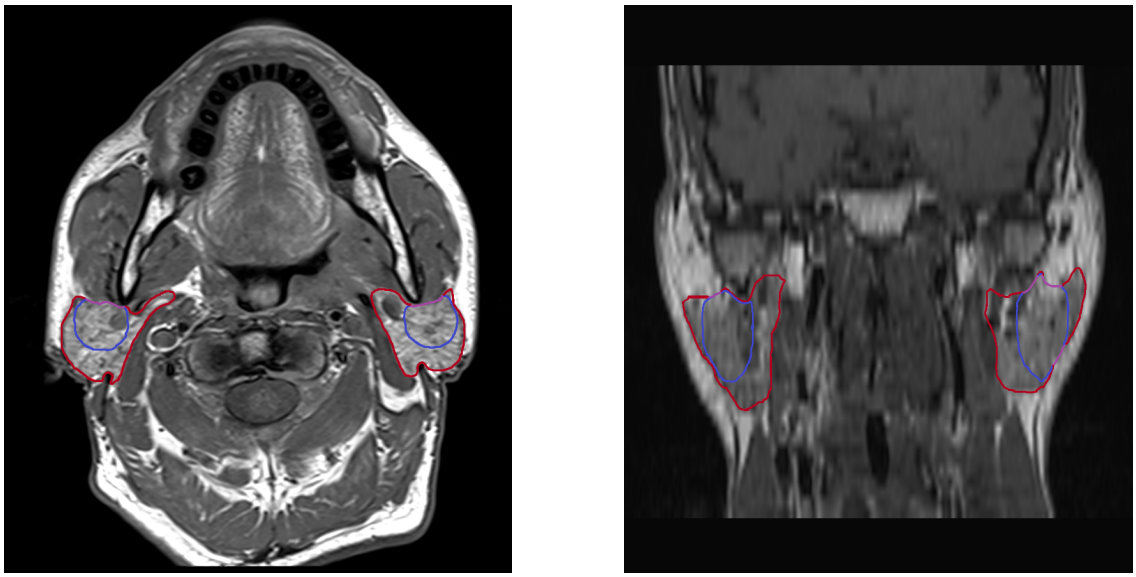
Intensity and textural IBMs of the SCR region were extracted by in-house software in MATLAB [83]. The determination of the IBMs is done according to guidelines created by the Image Biomarker Standardization Initiative (IBSI).[84] Textural IBMs can be divided in several classes, namely IBMs based on the grey-level co-occurence matrix (GLCM), IBMs based on the grey-level run length matrix (GLRLM) and IBMs based on the neighbourhood grey-tone difference matrix (NGTDM). The texture IBMs were extracted in 2D, and thus determined in only 4 directions. The determination of the intensity IBMs as well as the GLCM, GLRLM and NGTDM and their respective IBMs can be found in the Supplementary materials of *Zwanenburg et al. (2020)*.[84] In Appendix A, the used IBMs are stated. Morphologic IBMs were not calculated due to the geometrical determination of the SCR region, instead of an anatomical/intensity



(a) Original MRI scans.

(b) Normalized MRI scan.

Figure 1: Normalizing MRI according to the standardization of intensities technique. Example of two patients with varying intensities in original MRI scans.



(a) (b)

Figure 2: Parotid glands (red) and SCR regions (blue) in the transversal (a.) and coronal (b.) plane.

based determination.

2.2.5 Multiple imputation for missing values

The reason for data to be missing can have several causes, including treatment related causes (i.e., not being able to fill in a questionnaire due to severe complaints resulting from the treatment), which could introduce bias in the complete case scenario. Multiple imputation of missing data to be able to have a larger study population for predictive modelling has been proven to give less biased results than a complete case analysis, where missing data is excluded [85, 86]. Multiple imputation of missing clinical data was performed according to the multiple imputation methods of *van Buuren et al. (2011)* in R (version 4.0.5) [87]. In the multivariate imputation by chained equations (MICE) method, all other variables were used to estimate the values for the missing data. This has been executed 10 times. These results of analyses in each imputation set were pooled according to Rubin's rule [88].

2.2.6 Endpoint

The endpoint of this study was patient-rated daytime 12 months after treatment according to the Groningen Radiotherapy-Induced Xerostomia (GRIX) questionnaire. The full question, translated from Dutch to English, is 'Do you have a dry mouth during the day?' and this can be answered in a four point Likert-scale fashion (not at all vs. a bit vs. quite a bit vs. a lot). This endpoint is dichotomized as none-to-mild (including not at all and a bit) vs. moderate-to-severe (including quite a bit and a lot) for the purpose of this study.

2.2.7 Reference model

The clinical reference model was the NTCP model created by *van Rijn-Dekker et al. (2021)*[89]. This model was developed using multivariable logistic regression with moderate-to-severe daytime xerostomia as the endpoint. For this model, clinical parameters and dose parameters have been evaluated as possible predictor. The mean dose to the contralateral SCR region, the mean dose to the oral cavity and the pretreatment daytime xerostomia score (None vs. Any) were implemented in this model. Both before

and after external validation, this model had an AUC of 0.68 with an intercept and slope of -0.002 and 0.981 before and 0.254 and 0.965 after external validation respectively. The regression coefficients for this reference model can be found in Appendix B, Table A.2.

2.2.8 IBM model development for SCR region

Model development was executed in R (version 4.0.5). An univariable analysis was performed to investigate whether individual IBMs have a correlation higher than 0.8 with the endpoint discussed in section 2.2.6 *Endpoint*. Furthermore, in this univariable analysis, transformations of the variables were proposed which may have a higher correlation with the endpoint.

Multicollinearity, i.e. multiple predictors that are highly correlated to each other, can cause an unreliable and unstable estimate of regression coefficients. Therefore, a pre-selection of predictors was necessary, which was performed according to the methods of *van den Bosch et al. (2020)* [86]. As a first step, the IBMs with the best Bayesian Information Criterion (BIC) to predict the endpoint was found and included, after which all IBMS that had a Spearman correlation of >0.8 were excluded. This was done iteratively until all IBMs were either included or excluded. The second step in this pre-selection was to create predictor groups. In these predictor groups, each predictor had a Spearman correlation with the other predictors within the predictor group of less then 0.8, while making the predictor groups as large as possible. Henceforth, a logistic regression model was created based on each predictor group using forward selection based on the BIC. When a variable in the predictor group was selected by the model in more than 5 imputation sets, this variable was selected as a predictor. Hereafter, a logistic regression model was trained with these selected predictors. These models were then pooled according to Rubin's rules and these resulting models are called sub-models. To select the most relevant sub-models, a BIC threshold was defined by taking the lowest BIC along all sub-models and adding the natural logarithm of the amount of patients. All sub-models with a BIC equal to or lower than this threshold were selected for the final model, the composite model. For the composite model, the average of the regression coefficients of the linear predictors of the individual sub-models was taken.

The discriminating power of all models is evaluated with the AUC and a calibration curve. By using a Hosmer-Lemeshow test with 10 subgroups, the goodness of fit is analyzed. In a Hosmer-Lemeshow test, the development data is stratified in 10 groups according to their predicted outcome (all predicted values which are 0-10% of the maximum predicted value will be in group 1, for 10-20% of the maximum predicted value they will be in group 2, etc.). For these groups, the observed outcome is also determined. The mean of the observed and predicted values in these groups were depicted as dots in the calibration plot. If the outcome of the Hosmer-Lemeshow test is significant, this represents a lack-of-fit.

Hereafter, the model was internally validated to correct for optimism and prevent overfitting. This was executed by bootstrapping 100 times in each imputation set, so 1000 times in total. In this procedure, the amount of patients in the development cohort was selected with replacement out of the original patient group. Only IBMs of the predictor groups that were relevant in creating the composite model in the previous model development step were considered as possible predictors. Again, sub-models were created and all these sub-models were again pooled in a composite model. Eventually, it can be seen which IBMs were selected in each bootstrap sample. Optimism is defined as the difference between the performance of the model on the bootstrap sample and the performance of the model on the original patient group. Ultimately, the optimism was averaged over all bootstrap samples to find the optimism of the whole model. To remove this optimism from the model, the intercept and regression coefficients were updated.

2.2.9 Addition of IBMs to reference model

At first, the reference model was externally validated on our study population by evaluating the performance of the reference model with the original intercept and regression coefficients. Hereafter, a closed-testing procedure is performed to update the model to fit our study population better.[86] A closed-testing procedure indicates whether a model would fit a new study population better by adjusting the intercept, adjusting the intercept and slope or re-estimating all regression coefficients, which is the adjustment with the highest magnitude. As a first step, the original model is tested against a model with the highest magnitude of adjustment with a likelihood-ratio test. If the model with re-estimated regression coefficients

performs significantly better than the original model, it is concluded that a model adaptation needs to be applied. Hereafter, the model with re-estimated regression coefficients is tested against the model with only an adjusted intercept. If the model with an updated intercept is significantly better, this model is adopted. Otherwise, this process repeats with the model with re-estimated regression coefficients and the model with an updated intercept and slope. This process is illustrated in Figure 3.

Hereupon, the selected IBMs in the IBM-only model were added to the reference model. Using a likelihood-ratio test, it was examined whether addition of the IBMs improves results for predicting daytime xerostomia at 12 months after treatment.

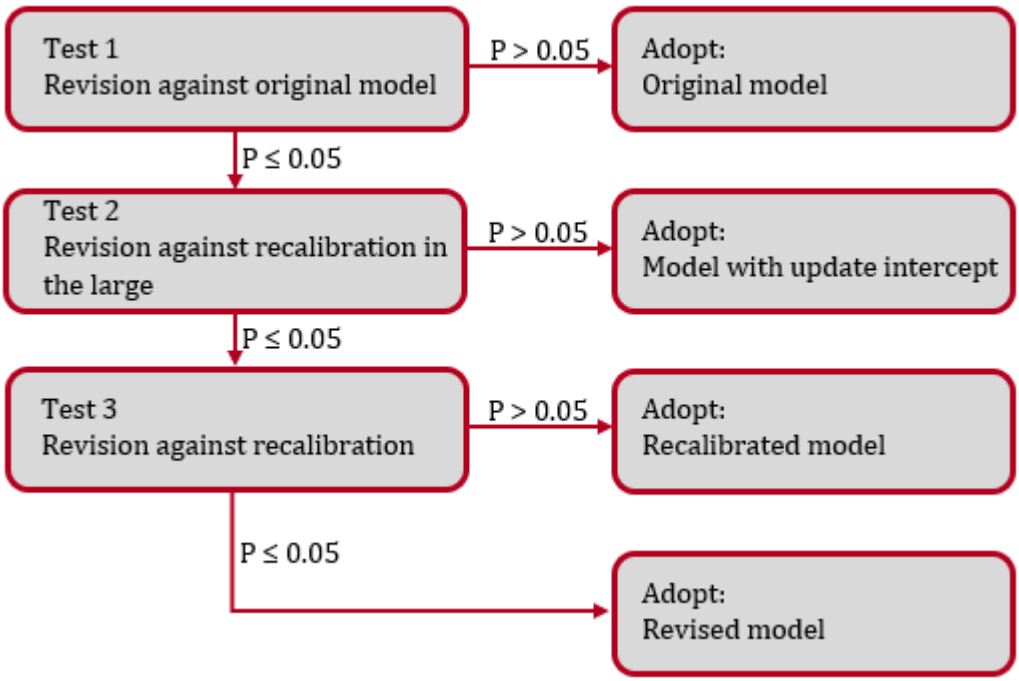


Figure 3: Closed-testing procedure. Figure is recreated based of figure in section 5.2 of the Supplementary Materials of *van den Bosch et al. (2020)*.[86]

2.2.10 Fit models IBMs whole parotid gland

Studies have been conducted which examine the influence of whole parotid gland IBMs on the prediction of xerostomia. The model of *van Dijk et al. (2018)* was created based on similar data as used in this study, but during a different time period.[59] To evaluate whether this already externally validated model also fits on our study population, IBMs were extracted for the whole parotid gland in the same manner as the IBMs for the SCR region and normalized by subtracting each value by the mean IBM value and hereupon dividing by the standard deviation of that IBM variable. Hereafter, the original model (see Appendix B Table A.3) with IBMs of the parotid gland and clinical parameters was externally validated on our study population. Henceforth, a closed-testing procedure is performed to update the model to fit our study population better.

2.3 Results

2.3.1 Study population demographics

Based on the flowchart in Appendix C, Figure A.1, 1519 patients with an MRI scan created for the Radiotherapy department were excluded. Ultimately, 104 patients were included. Due to time constraints, it

was not possible to delineate the contours on the MRI scan for 31 patients. In Table 2, the characteristics of the patients are stated. Most patients had large, locally advanced oropharyngeal tumours. The amount of missing data prior to imputation can be found in Appendix D, Table A.5. Only values considering daytime xerostomia were missing. All IBMs were determined based on the available MRI scans. After imputation of the endpoint, the distribution of the severity of the endpoint had changed, as can be seen in Appendix D, Figure A.2, panel M12. Out of all predictors, the most predictive parameters for whether the endpoint was missing or the value of the endpoint were answers to other xerostomia-related questions in the questionnaires at other time points and whether the patient had died. Prior to treatment, almost 50% of the patients experienced daytime xerostomia, of which 17% experienced moderate-to-severe daytime xerostomia. This increased to 79% and 38% respectively at 12 months after radiotherapy.

2.3.2 IBM-only model for SCR region

In the univariable analysis, all IBMs were examined on correlation with the daytime xerostomia endpoint. None of the correlations was proven to be significant.

In the multivariable analysis, two IBMs were selected, the short run high grey level emphasis (SRHGE) of the GLRLM of the contralateral SCR region and the long run low grey level emphasis (LRLGE) of the GLRLM of the ipsilateral SCR region. These variables were selected in 49 of the 96 sub-models that were created, the other sub-models did not contain any predictors since no IBMs were selected using the majority-rule. All sub-models where contralateral SRHGE and ipsilateral LRLGE were selected were identical, which in turn is also the result of the composite model. The intercept and regression coefficients can be found in Table 3. It was observed that only the SRHGE parameter was significantly selected in the procedure. The performance of the composite model before and after internal validation can be found in Table 4. The histogram of the IBM selection in the bootstrapping procedure of the internal validation can be found in Figure 5. It can be observed that the ipsilateral LRLGE was selected in 503 out of 1000 sub-models and the contralateral SRHGE was selected in 405 out of 1000 sub-models. The other IBMs were selected in less (69 to 249 out of 1000) sub-models. Before internal validation, the model had an AUC of 0.68 (95% CI 0.43-0.86), and after the internal validation, this was reduced to 0.57 (95% CI 0.45-0.68). Furthermore, the calibration plot of the composite model before and after internal validation can be found in Figure 4. Before internal validation, the groups created by the Hosmer-Lemeshow test along the calibration curve seem to lie along the curve itself, which is supported by the slope and intercept of Table 4. However, after internal validation, the groups created by the Hosmer-Lemeshow test lie further from the calibration curve and more grouped along a part of the predicted rate (0.30-0.45) instead of the full range. This indicates that this model predicts in a small range of the outcome, while the true range in the outcome is larger (0.14-0.81).

Table 2: Patient and dose characteristics.

Characteristics	**n = 104**
Gender	
Male	71
Female	33
Age	
Mean ± standard deviation (years)	60 ± 11
Tumour Location	
Larynx	3
Hypopharynx	1
Oropharynx	74
Nasopharynx	13
Oral cavity	9
Other	4
Tumour stage	
T1	23
T2	14
T3	17
T4	50
Nodal stage	
N0	22
N1	15
N2	62
N3	5
Treatment method	
Conventional radiotherapy	28
Accelerated radiotherapy	22
Conventional radiotherapy & chemotherapy	50
Accelerated radiotherapy & immunotherapy	4
Treatment modality	
Photons	43
Protons	61
Daytime xerostomia baseline	
Any xerostomia present*	49%
Xerostomia rated as moderate-to-severe*	17%
Daytime xerostomia M12	
Any xerostomia present*	79%
Xerostomia rated as moderate-to-severe*	38%
Dose to parotid gland	
Ipsilateral (median (IQR[§]))	25.46 Gy (19.87 Gy to 32.96 Gy)
Contralateral (median (IQR[§]))	15.60 Gy (8.87 Gy to 20.42 Gy)
Dose to SCR region	
Ipsilateral (median (IQR[§]))	20.62 Gy (13.48 Gy to 33.22 Gy)
Contralateral (median (IQR[§]))	9.09 Gy (5.19 Gy to 13.35 Gy)

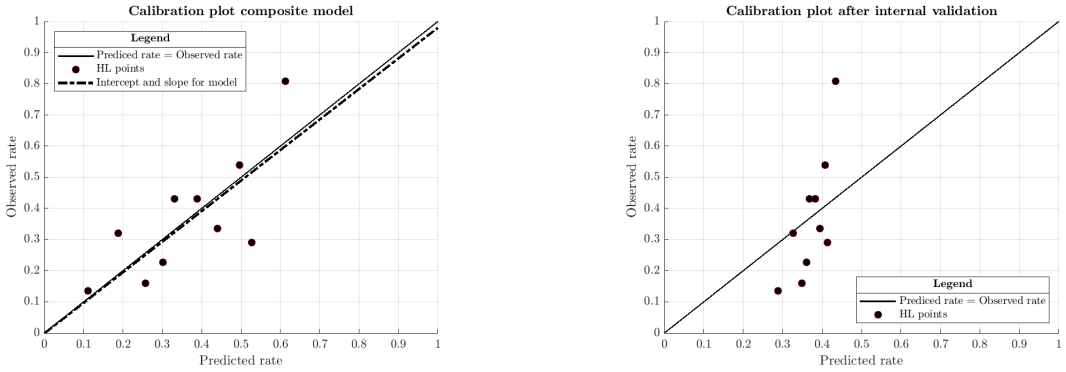
* Based on imputated data, see Appendix D, Table A.5 for amount of missing data in these variables.
§ IQR = inter quartile range.

Table 3: Regression coefficients from composite IBM-only model and IBM-only model after internal validation.

Model predictors	**Regression coefficient original**	**P-value**	**Regression coefficient after internal validation**
Intercept	3.08	0.03	0.36
LRLGE ipsilateral	-4.81	0.09	-1.15
SRHGE contralateral	-0.06	0.01	-0.01

Table 4: Performance composite IBM-only model

Model	AUC (95% CI)	Intercept	Slope
Composite model	0.68 (0.43-0.86)	-0.0014	0.98
Composite model internal validation	0.57 (0.45-0.68)	1.53	4.11



(a) (b)

Figure 4: Calibration plot IBMs-only model before (a.) and after (b.) internal validation.

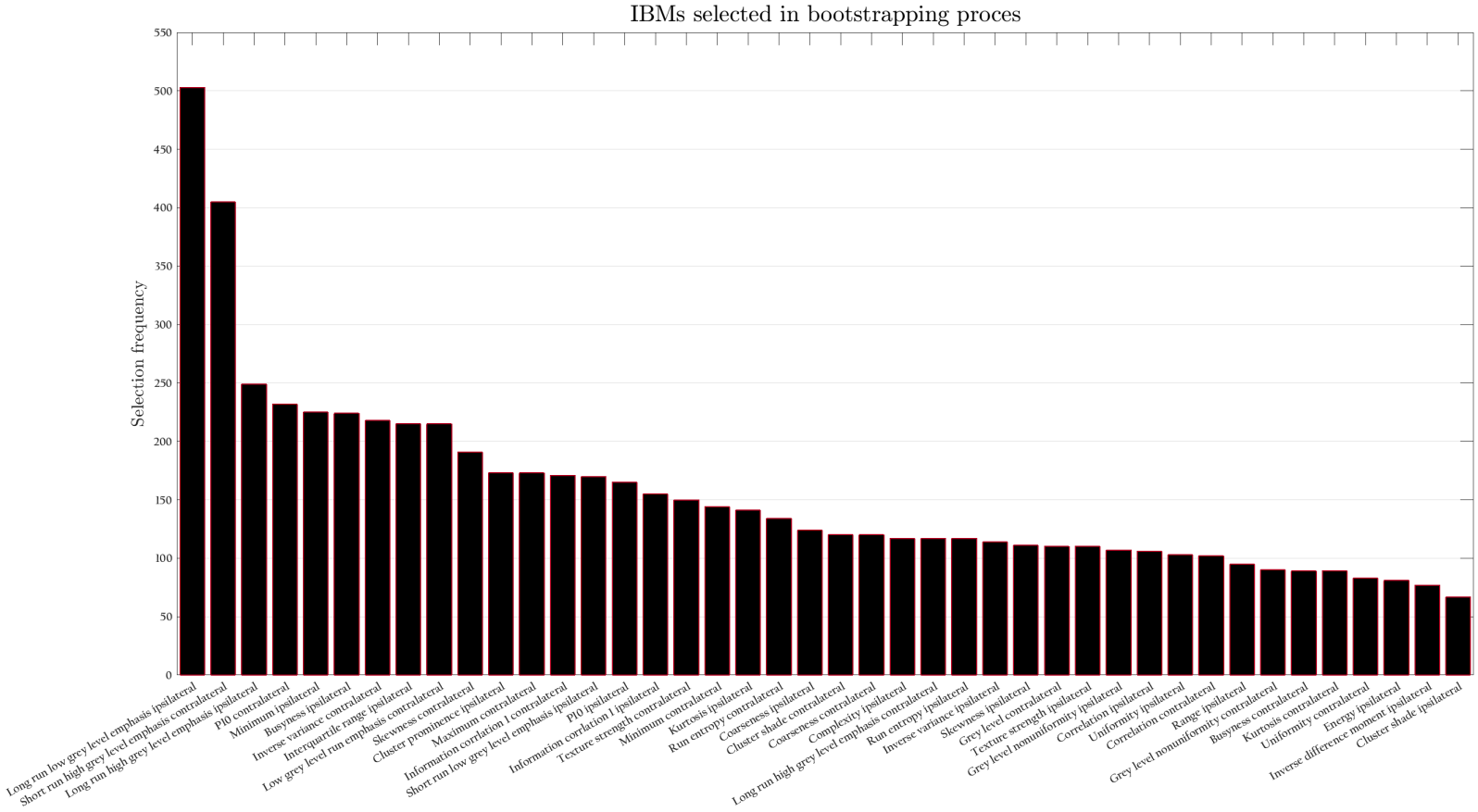


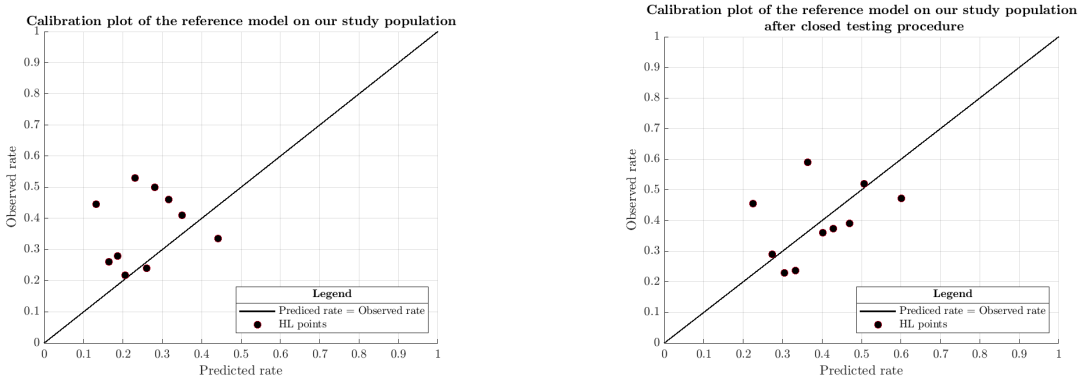
Figure 5: Selection frequency of all IBMs that were considered during internal validation. Total samples = 1000.

2.3.3 Addition IBMs to reference model

The performance of the reference model after the execution of the closed-testing procedure can be found in Table 5 and Figure 6. In the study in which the reference model was developed and externally validated, it had an AUC of 0.68, which was reduced to 0.55 on our study population. The updated intercept and regression coefficients of the reference model on our study population can be found in Appendix B, Table A.2. In the closed-testing procedure, it became apparent that a change in intercept would be the most beneficial for the model in 8 out of 10 imputation sets. In the other 2 sets, no change to the model was suggested. After the closed-testing procedure, the points of the Hosmer-Lemeshow test in the calibration plot lie more towards the line where the observed rate is equal to the predicted rate. Furthermore, the range in which the predictions have been done has become somewhat larger, while also giving a somewhat higher prediction rate (from 0.13-0.44 to 0.23-0.60). When the IBMs were added to the reference model and compared to the reference model alone with the likelihood-ratio, there was no significant improvement of the reference model (p = 0.998).

Table 5: Performance reference model on our study population.

Model	AUC	Intercept	Slope
Reference model on our study population	0.53	-0.43	0.10
Reference model after closed-testing procedure	0.55	-0.29	0.32



(a) (b)

Figure 6: (a.) Calibration plots of reference model on our study population. (b.) Calibration plots of reference model on our study population after recalibration according to closed-testing procedure.

2.3.4 Fit models IBMs whole parotid gland

The intensity and texture models created by *van Dijk et al. (2018)* were validated on our study population. In Figure 7a, the calibration plots of the intensity model (above) and the texture model (below) are depicted. The AUC, intercept and slope of these models can be found in Table 6. In the closed-testing procedure, it became apparent that a recalibration (i.e. updated intercept and slope) of both models gave a significantly better model on our data. The updated intercepts and regression coefficients can be found in Appendix B, table A.4. In the original publication by *van Dijk et al. (2018)*, the AUC of the externally validated intensity and texture models were both 0.83, whereas in our study, they were reduced to 0.69. As can be seen in the calibration curves, the recalibration of the model has caused the Hosmer-Lemeshow points to lie along the line where the observed rate is equal to the predicted rate. For the intensity model, a recalibration was suggested in 5 out of 10 imputation sets, a revision was suggested in 4 out of 10

imputation sets and the original model was suggested in 1 out of 10 imputation sets. For the texture model, a recalibration was suggested in 6 out of 10 imputation sets and a revision was suggested in 4 out of 10 imputation sets.

Table 6: Performance parotid gland IBM model on our study population.

Model	AUC	Intercept	Slope
Intensity			
Parotid gland IBM model on our study population	0.69	-0.01	0.50
Parotid gland IBM model after closed-testing procedure	0.69	0	1
Texture			
Parotid gland IBM model on our study population	0.67	0.51	0.57
Parotid gland IBM model after closed-testing procedure	0.69	0	1

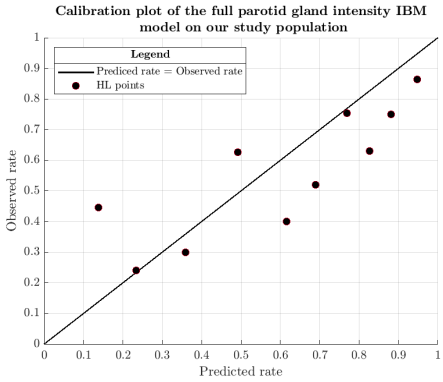
2.4 Discussion

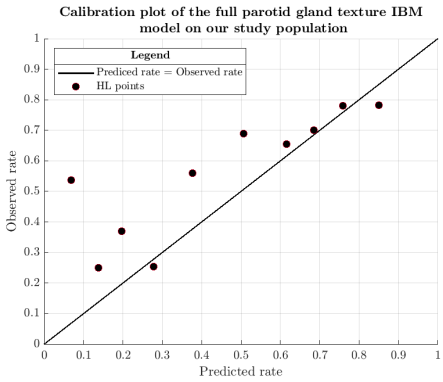
In the current prediction models for xerostomia, clinical parameters such as baseline xerostomia and dose to the parotid gland in combination with imaging parameters are used. With these models, the amount of xerostomia a patients will experience after radiotherapy can be predicted quite well. However, there is still room for improvement. The hypothesis that including IBMs of the SCR region of the parotid gland would also improve prediction models partially based on the SCR region was tested in this study. The results of this study did not find convincing benefits from adding IBMs from the SCR region to a reference model.

The IBMs that were selected for the model were both based on the GLRLM. The GLRLM is a measure for the repetition of the same intensity within a structure. An example of the calculation of the GLRLM can be found in the IBSI reference manual, page 84-85, which is part of the Supplementary Materials of the article by *Zwanenburg et al. (2020)*.[84] The LRLGE represents the amount of long runs of low intensities, where a run is a repetition of an amount of pixels with the same pixel value after binning. The SRHGE on the other hand represents the amount of short runs of high intensities. The LRLGE was selected for the ipsilateral SCR region, while the SRHGE was selected for the contralateral SCR region. Both have a protective effect for the development of xerostomia, as stated by the negative regression coefficients. Both IBMs are measures for the homogeneity; when the LRLGE is high, the structure is more homogeneous for low intensities, whereas when the SRHGE is high, the structure is less homogeneous for high intensities. These IBMs are not mutually exclusive, but are correlated for the same structure, see Appendix E, Figure A.3. This may explain why the SRHGE and LRLGE are only considered separately in each structure after the pre-selection of variables. Furthermore, both IBMs individually did not have a strong univariable correlation with the endpoint, but together they do have a correlation with the endpoint. Therefore, it can be said that these IBMs are mutually reinforcing in predicting daytime xerostomia.

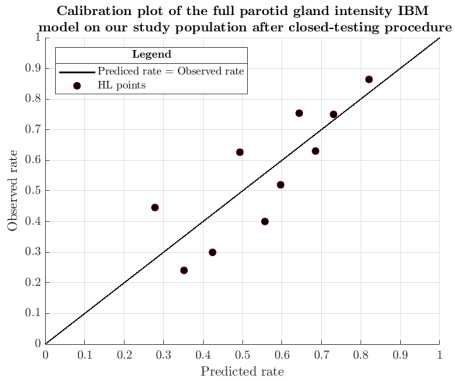
For the whole parotid and submandibular gland, IBMs indicating heterogeneity of tissue had a high predictive value for xerostomia. It was concluded that the amount of fatty tissue (non-functional tissue) in the salivary glands was a predictor for xerostomia.[59, 62] For the SCR region of the parotid gland, an IBM indicating homogeneity of low intensities and an IBM indicating heterogeneity of high intensities were found to be protective factors for daytime xerostomia at 12 months after radiotherapy. In essence, homogeneity of low intensities indicates functional tissue and the absence of homogeneity in high intensities would indicate that there are no large connecting parts of fatty tissue in the SCR region of the parotid gland. This brings us to a similar conclusion; homogeneity of parotid gland tissue and absence of large areas of fatty tissue are protective factors for developing daytime xerostomia at 12 months after radiotherapy.

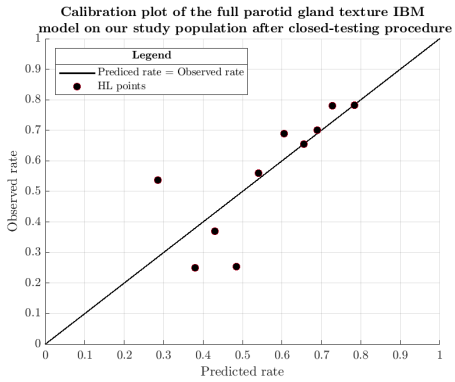
The performance of the model dropped severely after internal validation based on forward selection (AUC reduces from 0.68 to 0.57). In this procedure, 100 sub-models are created in each imputation set





(a) Calibration plots with original intercept and regression coefficients.





(b) Calibration plots with adjusted intercept and regression coefficients after closed-testing procedure.

Figure 7: Calibration plots of intensity (above) and texture (below) models by *van Dijk et al. (2018)* on our study population before (a.) and after (b.) closed-testing procedure.[59]

by a bootstrapping procedure, so 1000 sub-models in total are created. The LRLGE and SRHGE are selected in only 503 and 405 out of 1000 sub-models, respectively. The LRLGE, which was not statistically significant selected in the model-creation procedure (p = 0.09) is thus selected the most. Based on these outcomes, this decrease in performance can be accounted for. Furthermore, due to the pre-selection of predictors based on the BIC, some predictors are not taken into account during the internal validation. Currently, it was not possible to include all predictors due to the computational load this gave. In future research, it would be an addition to include all predictors during model development and internal validation, without pre-selection. As can be seen in Figure 4, the model has a smaller range in prediction rate after internal validation (Figure 4b) then before internal validation (Figure 4a). This indicates that the model was overfitted and thus optimistic before internal validation.

The clinical reference model by *van Rijn-Dekker et al. (2021)* did not perform well on our study population, even though the model was externally validated before and the performance did not change by that external validation (AUC remained 0.68)[89]. This indicates that the model is stable and can be applied to other external populations. After the closed-testing procedure, the AUC of the model on our study population was only 0.55. The likelihood-ratio test concluded that the addition of the IBM model did not improve the reference model. It was expected that the IBM model would improve the reference model, since the IBM model did have a better performance on this study population. However, the performance after internal validation was also reduced to 0.57, which was not much better than the reference model on this study population.

A comparison of predictors and endpoints in the reference model's population and our study population can be found in Appendix F, Table A.6. When these populations are compared, the dose parameters are significantly different, but the prevalence of xerostomia was not proven to be significantly different even though an increase is found. A difference in the study populations is the time during which the patients were included. For our study, this was between January 2018 and April 2020, while their inclusion period was March 2007 to January 2018. During the start of the inclusion period of their study and the end of the inclusion period of our study, many changes in planning techniques and treatment techniques have taken place. Due to the large cohort of their study, the study is more generalizable since more variation in data is proposed. Changes that have taken place were the introduction of Volumetric Arc Therapy (2014) and later Intensity Modulated Proton Therapy (2018), which both reduce the dose to organs at risk, such as the parotid gland and the oral cavity, while maintaining the dose to the tumour. [90, 91] Furthermore, there used to be a more prominent focus on sparing the parotid gland to prevent xerostomia, while the focus in later years turned to prevention of other side effects with a larger impact on the quality of life.[92] Differences in both dose and daytime xerostomia score at 12 months after radiotherapy can also be contributed to the larger percentage of prevalence of tumours to the larynx (39.8%) and hypopharynx (10.4%) they included. In our population, these prevalences were 2.8% and 0.9% respectively. The full parotid glands must be visible on MRI scans for our study, hence MRI scans with a focal point in the larynx or hypopharynx were usually excluded due to the small range of the MRI scan. As a result of the tumour location, patients with a larynx or hypopharynx carcinoma usually receive a smaller amount of dose to the parotid gland, resulting in a lower prevalence of daytime xerostomia at 12 months after radiotherapy.

As can be observed in Appendix D, Figure A.2, in most cases the distribution of the available data prior to imputation (red dot) does not lie within the boxplot, which represents the distribution of the data after imputation. As described previously, the reasons for missing data could introduce a certain bias. During imputation, this bias is corrected as best as possible. Due to this correction, it logically follows that the distribution of data prior to imputation is not in the area of the boxplot. In the available data prior to imputation, it can be observed that there is an overestimation of patients who do not have xerostomia, while there is an underestimation of patients who do experience xerostomia, except for the patients who experience severe (Very much in Figure A.2) xerostomia. What is most important in this figure, is that the distribution of complaints is attributable to clinical factors. During treatment (W01-W07 in Figure A.2), the prevalence of no xerostomia present diminishes, while the prevalence of xerostomia in each severity increases. These acute symptoms of radiotherapy are experienced by most patients, which is also an observation in the outpatient clinic. After treatment, the severity of xerostomia decreases until it hits a plateau in which most patients experience a stable state of xerostomia, which is the long-term effect of radiotherapy. In the outpatient clinic, it is observed that xerostomia slightly diminishes over time, especially in the first years after radiotherapy.

Multiple imputation of clinical data reduces bias. Furthermore, if only the complete cases where used for this study, the study population would be 40% smaller, which could compromise the ability to create a reliable model. Multiple imputation is a generally accepted and extensively tested method to deal with missing data, even with missing endpoints.[88, 93] Generally, imputation is accepted when the amount of missing data is less then 50% and can even be used with caution with up to 70% missing data. [94]

The parotid gland models did perform better in the original article by *van Dijk et al. (2018)* than in our study, as suggested by the AUC. This may have several causes. Again, the time period in which these patients were included differed, indicating that planning protocols and treatment techniques have changed, mainly affecting the dose parameter in the models. Again, it is also seen that the amount of laryngeal and hypopharyngeal cancer patients in their study is higher than in our study. Furthermore, the magnetic field strength in their study was 1.5T, whereas this is 3T in our study. A 3T scanner usually has a higher signal-to-noise ratio, which can influence both intensity and textural parameters. A comparison between the difference in parameters for these models in both study populations would also be advised to gain more insight in the reasons for the change in performance.

Due to time constraints, not all patients could be included. Transferring structures from CT scan to MRI scan was a more time-consuming process than expected in advance. Ultimately, structures were transferred for 104 patients, while there was no time left for the other 31 patients. To improve the model, it is recommended to add more patients for model development, which is likely to diversify the population and allow the model to take better consideration of specific cases. Furthermore, these patients could also be used for external validation of the model. Due to the small study population that was currently used for model development, it seemed more reasonable to use all of the patients to create the model rather than also using a portion for external validation.

The SCR region was defined geometrically based on knowledge from previous studies. Due to patient specific anatomical differences, the SCR region could be larger or smaller than this geometrically defined region. Furthermore, it is possible that other structures, like the external carotid artery and retromandibular vein, are present in the defined SCR region.[95] As described by *Pringle et al. (2013)* and tested by *van Luijk et al. (2016)*, stem cells of the parotid gland reside around the striated ducts.[40, 41] To ensure the selected region of interest only contains the region where stem cells reside, a method needs to be developed to visualize the striated ducts on imaging, after which a margin can be applied to ensure that the stem cells are included. MR sialography, in which the parotid ducts are visualized, may provide a solution. A specific heavily T2 weighted MR sequence, in which saliva acts as a contrast medium, is applied to visualize the ducts. These scans are created in a certain, non-standard position with a slight angulation to improve visualization.[96–98] As our study is retrospective, we were limited by the MRI scans that were available. The available MRI scans were created for delineation and dose planning purposes, not for sialography or specific imaging of the parotid gland. Therefore, the ductal system of the parotid gland was not visible on available T2 weighted scans. In some patients, the excretory duct (Stensen's duct) could be followed along its trajectory. Smaller ducts were not visible. To evaluate the role of the stem cell rich region in developing daytime xerostomia 12 months after radiotherapy better, modelling with IBMs of MR sialography images are expected to provide more conclusive results.

2.5 Conclusion

In the present study, a correlation was found between MRI IBMs of the parotid gland stem cell region and daytime xerostomia 12 months after radiotherapy. Nonetheless, these MRI IBMs did not add any additional predictive information to the clinical reference model. However, with a more precise definition of the SCR region, and a larger study population, MRI IBMs might have a higher predictive value for daytime xerostomia 12 months after radiotherapy in addition to the clinical parameters that were already presented.

References

[1] Christina Fitzmaurice et al. "Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-Adjusted life-years for 29 cancer groups, 1990 to 2017: A systematic analysis for the global burden of disease study". In: *JAMA Oncology* 5.12 (2019), pp. 1749–1768. ISSN: 23742445. DOI: 10.1001/jamaoncol.2019.2996.

[2] Integraal Kankercentrum Nederland. *Incidentie verschillende kankersoorten*. 2020. URL: https://www.iknl.nl/.

[3] Brandon A. Mahal et al. "Incidence and demographic burden of HPV-associated oropharyngeal head and neck cancers in the United States". In: *Cancer Epidemiology Biomarkers and Prevention* 28.10 (2019), pp. 1660–1667. ISSN: 10559965. DOI: 10.1158/1055-9965.EPI-19-0038.

[4] Ismail Matalka et al. "The incidence of Epstein-Barr virus in nasopharyngeal carcinoma of Jordanian patients". In: *European Archives of Oto-Rhino-Laryngology* 269.1 (2012), pp. 229–234. ISSN: 09374477. DOI: 10.1007/s00405-011-1562-6.

[5] cancer.gov. *Head and neck cancer fact sheet*. 2021. URL: https://www.cancer.gov/types/head-and-neck/head-neck-fact-sheet.

[6] Francesca Pezzuto et al. "Update on head and neck cancer: Current knowledge on epidemiology, risk factors, molecular features and novel therapies". In: *Oncology (Switzerland)* 89.3 (2015), pp. 125–136. ISSN: 14230232. DOI: 10.1159/000381717.

[7] Anne Aupérin. "Epidemiology of head and neck cancers: An update". In: *Current Opinion in Oncology* 32.3 (2020), pp. 178–186. ISSN: 1531703X. DOI: 10.1097/CCO.0000000000000629.

[8] Erich M. Sturgis, Qingyi Wei, and Margaret R. Spitz. "Descriptive epidemiology and risk factors for head and neck cancer". In: *Seminars in Oncology* 31.6 (2004), pp. 726–733. ISSN: 00937754. DOI: 10.1053/j.seminoncol.2004.09.013.

[9] Ruth Tachezy et al. "HPV and other risk factors of oral cavity / oropharyngeal cancer in the Czech Republic". In: *Oral Diseases* 11 (2005), pp. 181–185.

[10] Ana Lívia Silva Galbiatti et al. "Head and neck cancer: Causes, prevention and treatment". In: *Brazilian Journal of Otorhinolaryngology* 79.2 (2013), pp. 239–247. ISSN: 18088686. DOI: 10.5935/1808-8694.20130041. URL: http://dx.doi.org/10.5935/1808-8694.20130041.

[11] Shyh-An Yeh. "Radiotherapy for Head and Neck Cancer". In: *Seminars in Plastic Surgery* 24.02 (May 2010), pp. 127–136. ISSN: 1535-2188. DOI: 10.1055/s-0030-1255330. URL: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0030-1255330.

[12] Zhong He Wang et al. "Impact of salivary gland dosimetry on post-IMRT recovery of saliva output and xerostomia grade for head-and-neck cancer patients treated with or without contralateral submandibular gland sparing: A longitudinal study". In: *International Journal of Radiation Oncology Biology Physics* 81.5 (2011), pp. 1479–1487. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2010.07.1990.

[13] Michael F. Gensheimer et al. "Submandibular gland-sparing radiation therapy for locally advanced oropharyngeal squamous cell carcinoma: patterns of failure and xerostomia outcomes". In: *Radiation oncology (London, England)* 9 (2014), p. 255. ISSN: 1748717X. DOI: 10.1186/s13014-014-0255-x.

[14] Michael Little et al. "Reducing xerostomia after chemo-IMRT for head-and-neck cancer: Beyond sparing the parotid glands". In: *International Journal of Radiation Oncology Biology Physics* 83.3 (2012), pp. 1007–1014. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2011.09.004.

[15] Despina Katsochi. "Radiation Therapy with a Simultaneous Integrated Boost". In: *Radiotherapy*. InTech, May 2017. DOI: 10.5772/67326. URL: http://www.intechopen.com/books/radiotherapy/radiation-therapy-with-a-simultaneous-integrated-boost.

[16] Pierre Blanchard et al. "Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): A comprehensive analysis by tumour site". In: *Radiotherapy and Oncology* 100.1 (2011), pp. 33–40. ISSN: 01678140. DOI: 10.1016/j.radonc.2011.05.036. URL: http://dx.doi.org/10.1016/j.radonc.2011.05.036.

[17] Benjamin Lacas et al. "Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): An update on 107 randomized trials and 19,805 patients, on behalf of MACH-NC Group". In: *Radiotherapy and Oncology* 156 (2021), pp. 281–293. ISSN: 18790887. DOI: 10.1016/j.radonc.2021.01.013.

[18] Jean Bourhis et al. "Hyperfractionated or accelerated radiotherapy in head and neck cancer: a meta-analysis". In: *The Lancet* 368.9538 (Sept. 2006), pp. 843–854. ISSN: 01406736. DOI: 10.1016/S0140-6736(06)69121-6. URL: https://linkinghub.elsevier.com/retrieve/pii/S0140673606691216.

[19] Jens Overgaard et al. "Five compared with six fractions per week of conventional radiotherapy of squamous-cell carcinoma of head and neck: DAHANCA 6&7 randomised controlled trial". In: *The Lancet* 362.9388 (Sept. 2003), pp. 933–940. ISSN: 01406736. DOI: 10.1016/S0140-6736(03)14361-9. URL: https://linkinghub.elsevier.com/retrieve/pii/S0140673603143619.

[20] Benjamin Lacas et al. "Role of radiotherapy fractionation in head and neck cancers (MARCH): an updated meta-analysis". In: *The Lancet Oncology* 18.9 (Sept. 2017), pp. 1221–1237. ISSN: 14702045. DOI: 10.1016/S1470-2045(17)30458-8. URL: https://linkinghub.elsevier.com/retrieve/pii/S1470204517304588.

[21] Jean Bourhis et al. "Concomitant chemoradiotherapy versus acceleration of radiotherapy with or without concomitant chemotherapy in locally advanced head and neck carcinoma (GORTEC 99-02): an open-label phase 3 randomised trial". In: *The Lancet Oncology* 13.2 (Feb. 2012), pp. 145–153. ISSN: 14702045. DOI: 10.1016/S1470-2045(11)70346-1. URL: https://linkinghub.elsevier.com/retrieve/pii/S1470204511703461.

[22] Phuc Felix Nguyen-Tan et al. "Randomized Phase III Trial to Test Accelerated Versus Standard Fractionation in Combination With Concurrent Cisplatin for Head and Neck Carcinomas in the Radiation Therapy Oncology Group 0129 Trial: Long-Term Report of Efficacy and Toxicity". In: *Journal of Clinical Oncology* 32.34 (Dec. 2014), pp. 3858–3867. ISSN: 0732-183X. DOI: 10.1200/JCO.2014.55.3925. URL: http://ascopubs.org/doi/10.1200/JCO.2014.55.3925.

[23] James A Bonner et al. "Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival". In: *The Lancet Oncology* 11.1 (Jan. 2010), pp. 21–28. ISSN: 14702045. DOI: 10.1016/S1470-2045(09)70311-0. URL: https://linkinghub.elsevier.com/retrieve/pii/S1470204509703110.

[24] Maura L Gillison et al. "Radiotherapy plus cetuximab or cisplatin in human papillomavirus-positive oropharyngeal cancer (NRG Oncology RTOG 1016): a randomised, multicentre, non-inferiority trial". In: *The Lancet* 393.10166 (Jan. 2019), pp. 40–50. ISSN: 01406736. DOI: 10.1016/S0140-6736(18)32779-X. URL: https://linkinghub.elsevier.com/retrieve/pii/S014067361832779X.

[25] Hisham Mehanna 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". In: *The Lancet* 393.10166 (Jan. 2019), pp. 51–60. ISSN: 01406736. DOI: 10.1016/S0140-6736(18)32752-1. URL: https://linkinghub.elsevier.com/retrieve/pii/S0140673618327521.

[26] Ivo Beetz et al. "NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: The role of dosimetric and clinical factors". In: *Radiotherapy and Oncology* 105.1 (2012), pp. 101–106. ISSN: 01678140. DOI: 10.1016/j.radonc.2012.03.004. URL: http://dx.doi.org/10.1016/j.radonc.2012.03.004.

[27] Johannes A. Langendijk et al. "Selection of patients for radiotherapy with protons aiming at reduction of side effects: The model-based approach". In: *Radiotherapy and Oncology* 107.3 (2013), pp. 267–273. ISSN: 18790887. DOI: 10.1016/j.radonc.2013.05.007.

[28] Johannes A. Langendijk et al. "National Protocol for Model-Based Selection for Proton Therapy in Head and Neck Cancer". In: *International Journal of Particle Therapy* 8.1 (2021), pp. 354–365. ISSN: 2331-5180. DOI: 10.14338/ijpt-20-00089.1.

[29] Zhong He Wang et al. "Radiation-induced volume changes in parotid and submandibular glands in patients with head and neck cancer receiving postoperative radiotherapy: A longitudinal study". In: *Laryngoscope* 119.10 (2009), pp. 1966–1974. ISSN: 0023852X. DOI: 10.1002/lary.20601.

[30] Sara Broggi et al. "A two-variable linear model of parotid shrinkage during IMRT for head and neck cancer". In: *Radiotherapy and Oncology* 94.2 (2010), pp. 206–212. ISSN: 01678140. DOI: 10.1016/j.radonc.2009.12.014. URL: http://dx.doi.org/10.1016/j.radonc.2009.12.014.

[31] Choonik Lee et al. "Evaluation of geometric changes of parotid glands during head and neck cancer radiotherapy using daily MVCT and automatic deformable registration". In: *Radiotherapy and Oncology* 89.1 (2008), pp. 81–88. ISSN: 01678140. DOI: 10.1016/j.radonc.2008.07.006.

[32] Eliana M. Vásquez Osorio et al. "Local Anatomic Changes in Parotid and Submandibular Glands During Radiotherapy for Oropharynx Cancer and Correlation With Dose, Studied in Detail With Nonrigid Registration". In: *International Journal of Radiation Oncology Biology Physics* 70.3 (2008), pp. 875–882. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2007.10.063.

[33] Chunhui Han et al. "Actual Dose Variation of Parotid Glands and Spinal Cord for Nasopharyngeal Cancer Patients During Radiotherapy". In: *International Journal of Radiation Oncology Biology Physics* 70.4 (2008), pp. 1256–1262. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2007.10.067.

[34] Jerry L. Barker et al. "Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system". In: *International Journal of Radiation Oncology Biology Physics* 59.4 (2004), pp. 960–970. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2003.12.024.

[35] Eric K. Hansen et al. "Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer". In: *International Journal of Radiation Oncology Biology Physics* 64.2 (2006), pp. 355–362. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2005.07.957.

[36] James L. Robar et al. "Spatial and Dosimetric Variability of Organs at Risk in Head-and-Neck Intensity-Modulated Radiotherapy". In: *International Journal of Radiation Oncology Biology Physics* 68.4 (2007), pp. 1121–1130. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2007.01.030.

[37] Kimberly J. Jasmer et al. "Radiation-Induced Salivary Gland Dysfunction: Mechanisms, Therapeutics and Future Directions". In: *Journal of Clinical Medicine* 9.12 (2020), p. 4095. ISSN: 2077-0383. DOI: 10.3390/jcm9124095.

[38] Pètra M. Braam et al. "Long-term parotid gland function after radiotherapy". In: *International Journal of Radiation Oncology Biology Physics* 62.3 (2005), pp. 659–664. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2004.12.015.

[39] Antonius W.T. Konings, Rob P. Coppes, and Arjan Vissink. "On the mechanism of salivary gland radiosensitivity". In: *International Journal of Radiation Oncology Biology Physics* 62.4 (2005), pp. 1187–1194. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2004.12.051.

[40] Sarah Pringle, Ronald Van Os, and Robert P. Coppes. "Concise review: Adult salivary gland stem cells and a potential therapy for xerostomia". In: *Stem Cells* 31.4 (2013), pp. 613–619. ISSN: 10665099. DOI: 10.1002/stem.1327.

[41] Peter van Luijk et al. "Sparing the region of the salivary gland containing stem cells preserves saliva production after radiotherapy for head and neck cancer". In: *Science Translational Medicine* 7.305 (Sept. 2015), pp. 147–305. ISSN: 1946-6234. DOI: 10.1126/scitranslmed.aac4441. URL: https://stm.sciencemag.org/lookup/doi/10.1126/scitranslmed.aac4441.

[42] Antonius W.T. Konings et al. "Volume effects and region-dependent radiosensitivity of the parotid gland". In: *International Journal of Radiation Oncology Biology Physics* 62.4 (2005), pp. 1090–1095. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2004.12.035.

[43] Avraham Eisbruch et al. "Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer". In: *International Journal of Radiation Oncology Biology Physics* 50.3 (2001), pp. 695–704. ISSN: 03603016. DOI: 10.1016/S0360-3016(01)01512-7.

[44] Peter G. Hawkins et al. "Sparing all salivary glands with IMRT for head and neck cancer: Longitudinal study of patient-reported xerostomia and head-and-neck quality of life". In: *Radiotherapy and Oncology* 126.1 (2018), pp. 68–74. ISSN: 18790887. DOI: 10.1016/j.radonc.2017.08.002. URL: http://dx.doi.org/10.1016/j.radonc.2017.08.002.

[45] Ivo Beetz et al. "The QUANTEC criteria for parotid gland dose and their efficacy to prevent moderate to severe patient-rated xerostomia". In: *Acta Oncologica* 53.5 (2014), pp. 597–604. ISSN: 1651226X. DOI: 10.3109/0284186X.2013.831186.

[46] Volker Rudat et al. "The effect of amifostine or IMRT to preserve the parotid function after radiotherapy of the head and neck region measured by quantitative salivary gland scintigraphy". In: *Radiotherapy and Oncology* 89.1 (Oct. 2008), pp. 71–80. ISSN: 01678140. DOI: 10.1016/j.radonc.2008.07.016. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167814008003812.

[47] Bhadrasain Vikram and David M. Brizel. "Phase III Randomized Trial of Amifostine as a Radioprotector in Head and Neck Cancer". In: *Journal of Clinical Oncology* 19.4 (Feb. 2001), pp. 1233–1234. ISSN: 0732-183X. DOI: 10.1200/JCO.2001.19.4.1233. URL: http://ascopubs.org/doi/10.1200/JCO.2001.19.4.1233.

[48] Jens Buentzel et al. "Intravenous amifostine during chemoradiotherapy for head-and-neck cancer: A randomized placebo-controlled phase III study". In: *International Journal of Radiation Oncology Biology Physics* 64.3 (2006), pp. 684–691. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2005.08.005.

[49] Maverick GL Lee et al. "Randomized double-blind trial of amifostine versus placebo for radiation-induced xerostomia in patients with head and neck cancer". In: *Journal of Medical Imaging and Radiation Oncology* 63.1 (Feb. 2019), pp. 142–150. ISSN: 1754-9477. DOI: 10.1111/1754-9485.12833. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1754-9485.12833.

[50] Kristy E. Gilman et al. "P2X7 receptor deletion suppresses γ-radiation-induced hyposalivation". In: *American Journal of Physiology - Regulatory Integrative and Comparative Physiology* 316.5 (2019), R687–R696. ISSN: 15221490. DOI: 10.1152/ajpregu.00192.2018.

[51] Sten M. Wie et al. "Tyrosine Kinase Inhibitors Protect the Salivary Gland from Radiation Damage by Inhibiting Activation of Protein Kinase C-δ". In: *Molecular Cancer Therapeutics* 16.9 (Sept. 2017), pp. 1989–1998. ISSN: 1535-7163. DOI: 10.1158/1535-7163.MCT-17-0267. URL: http://mct.aacrjournals.org/lookup/doi/10.1158/1535-7163.MCT-17-0267.

[52] Maria Morgan-Bathke et al. "The rapalogue, CCI-779, improves salivary gland function following radiation". In: *PLoS ONE* 9.12 (2014), pp. 1–22. ISSN: 19326203. DOI: 10.1371/journal.pone.0113183.

[53] Zhao Zhu et al. "Prevention of irradiation-induced salivary hypofunction by rapamycin in swine parotid glands". In: *Oncotarget* 7.15 (2016), pp. 20271–20281. ISSN: 19492553. DOI: 10.18632/oncotarget.7941.

[54] Grace Hill et al. "Pharmacological activation of the EDA/EDAR signaling pathway restores salivary gland function following radiation-induced damage". In: *PLoS ONE* 9.11 (2014). ISSN: 19326203. DOI: 10.1371/journal.pone.0112840.

[55] Roel J.H.M. Steenbakkers. *Preventing Head and Neck Cancer Patients from developing permanent Radiation induced Xerostomia: The X-PREVENT Project*. 2017.

[56] Roel J.H.M. Steenbakkers et al. "Parotid gland stem cell sparing radiotherapy for head and neck cancer patients: a double-blind randomized controlled trial". In: *International Journal of Radiation Oncology*Biology*Physics* (2021). ISSN: 03603016. DOI: 10.1016/j.ijrobp.2021.09.023. URL: https://doi.org/10.1016/j.ijrobp.2021.09.023.

[57] Lisanne V. van Dijk et al. "Geometric Image Biomarker Changes of the Parotid Gland Are Associated With Late Xerostomia". In: *International Journal of Radiation Oncology Biology Physics* 99.5 (2017), pp. 1101–1110. ISSN: 1879355X. DOI: 10.1016/j.ijrobp.2017.08.003. URL: https://doi.org/10.1016/j.ijrobp.2017.08.003.

[58] Lisanne V. van Dijk et al. "CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva". In: *Radiotherapy and Oncology* 122.2 (2017), pp. 185–191. ISSN: 18790887. DOI: 10.1016/j.radonc.2016.07.007. URL: http://dx.doi.org/10.1016/j.radonc.2016.07.007.

[59] Lisanne V. van Dijk et al. "Parotid gland fat related Magnetic Resonance image biomarkers improve prediction of late radiation-induced xerostomia". In: *Radiotherapy and Oncology* 128.3 (2018), pp. 459–466. ISSN: 18790887. DOI: 10.1016/j.radonc.2018.06.012. URL: https://doi.org/10.1016/j.radonc.2018.06.012.

[60] Lisanne V. van Dijk et al. "18F-FDG PET image biomarkers improve prediction of late radiation-induced xerostomia". In: *Radiotherapy and Oncology* 126.1 (2018), pp. 89–95. ISSN: 18790887. DOI: 10.1016/j.radonc.2017.08.024. URL: https://doi.org/10.1016/j.radonc.2017.08.024.

[61] Lisanne V. van Dijk et al. "Delta-radiomics features during radiotherapy improve the prediction of late xerostomia". In: *Scientific Reports* 9.1 (2019), pp. 1–8. ISSN: 20452322. DOI: 10.1038/s41598-019-48184-3. URL: http://dx.doi.org/10.1038/s41598-019-48184-3.

[62] Khadija Sheikh et al. "Predicting acute radiation induced xerostomia in head and neck Cancer using MR and CT Radiomics of parotid and submandibular glands". In: *Radiation Oncology* 14.1 (2019), pp. 1–11. ISSN: 1748717X. DOI: 10.1186/s13014-019-1339-4.

[63] Minoru Nakatsugawa et al. "Radiomic Analysis of Salivary Glands and Its Role for Predicting Xerostomia in Irradiated Head and Neck Cancer Patients". In: *International Journal of Radiation Oncology*Biology*Physics* 96.2 (2016), S217. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2016.06.539. URL: http://dx.doi.org/10.1016/j.ijrobp.2016.06.539.

[64] Valerio Nardone et al. "Texture analysis as a predictor of radiation-induced xerostomia in head and neck patients undergoing IMRT". In: *Radiologia Medica* 123.6 (2018), pp. 415–423. ISSN: 18266983. DOI: 10.1007/s11547-017-0850-7. URL: https://doi.org/10.1007/s11547-017-0850-7.

[65] Benjamin S Rosen et al. "Early Changes in Serial CBCT-Measured Parotid Gland Biomarkers Predict Chronic Xerostomia After Head and Neck Radiation Therapy". In: *International Journal of Radiation Oncology Biology Physics* 102.4 (Nov. 2018), pp. 1319–1329. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2018.06.048. URL: https://linkinghub.elsevier.com/retrieve/pii/S0360301618310496.

[66] Elisa Scalco et al. "Texture analysis for the assessment of structural changes in parotid glands induced by radiotherapy". In: *Radiotherapy and Oncology* 109.3 (2013), pp. 384–387. ISSN: 01678140. DOI: 10.1016/j.radonc.2013.09.019. URL: http://dx.doi.org/10.1016/j.radonc.2013.09.019.

[67] Hui Wu et al. "Early Prediction of Acute Xerostomia During Radiation Therapy for Head and Neck Cancer Based on Texture Analysis of Daily CT". In: *International Journal of Radiation Oncology Biology Physics* 102.4 (2018), pp. 1308–1318. ISSN: 1879355X. DOI: 10.1016/j.ijrobp.2018.04.059. URL: https://doi.org/10.1016/j.ijrobp.2018.04.059.

[68] Masahiro Izumi et al. "MR imaging of the parotid gland in Sjögren's syndrome: a proposal for new diagnostic criteria." In: *American Journal of Roentgenology* 166.6 (June 1996), pp. 1483–1487. ISSN: 0361-803X. DOI: 10.2214/ajr.166.6.8633469. URL: http://www.ajronline.org/doi/10.2214/ajr.166.6.8633469.

[69] Karthik Ramasubramanian and Abhishek Singh. "The problem of multicollinearity". In: *Understanding Regression Analysis*. Berkeley, CA: Apress, 1997. Chap. 37, pp. 176–180. ISBN: 978-0-585-25657-3. DOI: 10.1007/978-0-585-25657-3{_}37. URL: http://link.springer.com/10.1007/978-0-585-25657-3_37.

[70] Louis B. Harrison et al. "Detailed quality of life assessment in patients treated with primary radiotherapy for squamous cell cancer of the base of the tongue". In: *Head and Neck* 19.3 (1997), pp. 169–175. ISSN: 01486403. DOI: 10.1002/(sici)1097-0347(199705)19:3<169::aid-hed1>3.0.co;2-0.

[71] Anke Petra Jellema et al. "Impact of Radiation-Induced Xerostomia on Quality of Life After Primary Radiotherapy Among Patients With Head and Neck Cancer". In: *International Journal of Radiation Oncology Biology Physics* 69.3 (2007), pp. 751–760. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2007.04.021.

[72] Anders B. Jensen et al. "Influence of late side-effects upon daily life after radiotherapy for laryngeal and pharyngeal cancer". In: *Acta Oncologica* 33.5 (1994), pp. 487–491. ISSN: 0284186X. DOI: 10.3109/02841869409083923.

[73] Marco Pota et al. "Early prediction of radiotherapy-induced parotid shrinkage and toxicity based on CT radiomics and fuzzy classification". In: *Artificial Intelligence in Medicine* 81 (2017), pp. 41–53. ISSN: 18732860. DOI: 10.1016/j.artmed.2017.03.004. URL: http://dx.doi.org/10.1016/j.artmed.2017.03.004.

[74] Cheng Jian Xu et al. "Impact of statistical learning methods on the predictive power of multivariate normal tissue complication probability models". In: *International Journal of Radiation Oncology Biology Physics* 82.4 (2012), e677–e684. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2011.09.036. URL: http://dx.doi.org/10.1016/j.ijrobp.2011.09.036.

[75] Antonetta C. Houweling et al. "A Comparison of Dose-Response Models for the Parotid Gland in a Large Group of Head-and-Neck Cancer Patients". In: *International Journal of Radiation Oncology Biology Physics* 76.4 (2010), pp. 1259–1265. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2009.07.1685.

[76] Ivo Beetz et al. "Development of NTCP models for head and neck cancer patients treated with three-dimensional conformal radiotherapy for xerostomia and sticky saliva: The role of dosimetric and clinical factors". In: *Radiotherapy and Oncology* 105.1 (2012), pp. 86–93. ISSN: 01678140. DOI: 10.1016/j.radonc.2011.05.010. URL: http://dx.doi.org/10.1016/j.radonc.2011.05.010.

[77] Hubert S. Gabryś et al. "Design and selection of machine learning methods using radiomics and dosiomics for normal tissue complication probability modeling of xerostomia". In: *Frontiers in Oncology* 8.MAR (2018), pp. 1–20. ISSN: 2234943X. DOI: 10.3389/fonc.2018.00035.

[78] Robert J. Gillies, Paul E. Kinahan, and Hedvig Hricak. "Radiomics: Images are more than pictures, they are data". In: *Radiology* 278.2 (2016), pp. 563–577. ISSN: 15271315. DOI: 10.1148/radiol.2015151169.

[79] Aaron D. Fain and Francesca D. Beaman. *Magnetic Resonance Imaging of Soft Tissue Masses*. Vol. 52. 4. Elsevier, 2017, pp. 227–240. ISBN: 8593235069. DOI: 10.1053/j.ro.2017.04.003. URL: http://dx.doi.org/10.1053/j.ro.2017.04.003.

[80] Andrés Larroza, Vicente Bodí, and David Moratal. "Texture Analysis in Magnetic Resonance Imaging: Review and Considerations for Future Applications". In: *Assessment of Cellular and Organ Function and Dysfunction using Direct and Derived MRI Methodologies*. InTech, Oct. 2016. DOI: 10.5772/64641. URL: http://www.intechopen.com/books/assessment-of-cellular-and-organ-function-and-dysfunction-using-direct-and-derived-mri-methodologies/texture-analysis-in-magnetic-resonance-imaging-review-and-considerations-for-future-applications.

[81] Nicolas Robitaille et al. "Tissue-based MRI intensity standardization: Application to multicentric datasets". In: *International Journal of Biomedical Imaging* 2012 (2012). ISSN: 16874188. DOI: 10.1155/2012/347120.

[82] Charlotte L. Brouwer et al. "CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines". In: *Radiotherapy and Oncology* 117.1 (2015), pp. 83–90. ISSN: 18790887. DOI: 10.1016/j.radonc.2015.07.041. URL: http://dx.doi.org/10.1016/j.radonc.2015.07.041.

[83] Lisanne V. van Dijk. *MATLAB scripts for determining IBMs*. 2021.

[84] Alex Zwanenburg et al. "The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping". In: *Radiology* 295.2 (2020), pp. 328–338. ISSN: 15271315. DOI: 10.1148/radiol.2020191145.

[85] Kristel J.M. Janssen et al. "Missing covariate data in medical research: To impute is better than to ignore". In: *Journal of Clinical Epidemiology* 63.7 (2010), pp. 721–727. ISSN: 08954356. DOI: 10.1016/j.jclinepi.2009.12.008. URL: http://dx.doi.org/10.1016/j.jclinepi.2009.12.008.

[86] Lisa Van den Bosch et al. "Key challenges in normal tissue complication probability model development and validation: towards a comprehensive strategy". In: *Radiotherapy and Oncology* 148 (2020), pp. 151–156. ISSN: 18790887. DOI: 10.1016/j.radonc.2020.04.012. URL: https://doi.org/10.1016/j.radonc.2020.04.012.

[87] Stef van Buuren and Karin Groothuis-Oudshoorn. "mice: Multivariate imputation by chained equations in R". In: *Journal of Statistical Software* 45.3 (2011), pp. 1–67. ISSN: 15487660. DOI: 10.18637/jss.v045.i03.

[88] Donald B. Rubin. *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, Ltd, 1987.

[89] Irene van Rijn-Dekker et al. "OC-0517 Importance of dose to the parotid gland stem cell rich region in preventing xerostomia in RT for HNC". In: *Radiotherapy and Oncology* 161 (Aug. 2021), S401–S402. ISSN: 01678140. DOI: 10.1016/S0167-8140(21)06943-7. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167814021069437.

[90] Andrea Holt et al. "Multi-institutional comparison of volumetric modulated arc therapy vs. intensity-modulated radiation therapy for head-and-neck cancer: A planning study". In: *Radiation Oncology* 8.1 (2013), p. 1. ISSN: 1748717X. DOI: 10.1186/1748-717X-8-26. URL: Radiation%20Oncology.

[91] Carina Behrends et al. “Model-based comparison of organ at risk protection between VMAT and robustly optimised IMPT plans”. In: *Zeitschrift fur Medizinische Physik* 31.1 (2021), pp. 5–15. ISSN: 18764436. DOI: 10.1016/j.zemedi.2020.09.003. URL: https://doi.org/10.1016/j.zemedi.2020.09.003.

[92] Hans Paul van der Laan et al. “Impact of radiation-induced toxicities on quality of life of patients treated for head and neck cancer”. In: *Radiotherapy and Oncology* 160 (2021), pp. 47–53. ISSN: 18790887. DOI: 10.1016/j.radonc.2021.04.011. URL: https://doi.org/10.1016/j.radonc.2021.04.011.

[93] Jared S. Murray. “Multiple Imputation: A Review of Practical and Theoretical Findings”. In: *Statistical Science* 33.2 (May 2018). ISSN: 0883-4237. DOI: 10.1214/18-STS644. URL: https://projecteuclid.org/journals/statistical-science/volume-33/issue-2/Multiple-Imputation-A-Review-of-Practical-and-Theoretical-Findings/10.1214/18-STS644.full.

[94] Ian R. White, Patrick Royston, and Angela M. Wood. “Multiple imputation using chained equations: Issues and guidance for practice”. In: *Statistics in Medicine* 30.4 (Feb. 2011), pp. 377–399. ISSN: 02776715. DOI: 10.1002/sim.4067. URL: https://onlinelibrary.wiley.com/doi/10.1002/sim.4067.

[95] Giuseppe Cicero et al. “Cross-sectional imaging of parotid gland nodules: A brief practical guide”. In: *Journal of Clinical Imaging Science* 8.1 (2018), pp. 1–8. ISSN: 21565597. DOI: 10.4103/jcis.JCIS.

[96] Eleftheria Astreinidou et al. “3D MR sialography protocol for postradiotherapy follow-up of the salivary duct system”. In: *Journal of Magnetic Resonance Imaging* 24.3 (2006), pp. 556–562. ISSN: 10531807. DOI: 10.1002/jmri.20659.

[97] Ankur Gadodia et al. “Magnetic resonance sialography using CISS and HASTE sequences in inflammatory salivary gland diseases: Comparison with digital sialography”. In: *Acta Radiologica* 51.2 (2010), pp. 156–163. ISSN: 02841851. DOI: 10.3109/02841850903376306.

[98] Nezahat Karaca Erdoğan et al. “Magnetic resonance sialography findings of submandibular ducts imaging”. In: *BioMed Research International* 2013 (2013). ISSN: 23146133. DOI: 10.1155/2013/417052.

Appendices

A IBMs

Table A.1: Used IBMs for model creation. IBMs are determined according to IBSI standards. [84]

Category	**IBM**	**Category**	**IBM**
Intensity	P10	GLCM	Joint average
Intensity	P90	GLCM	Joint entropy
Intensity	Entropy	GLCM	Joint maximum
Intensity	Interquartile range	GLCM	Joint variance
Intensity	Kurtosis	GLCM	Sum average
Intensity	Maximum	GLCM	Sum entropy
Intensity	Mean	GLCM	Sum variance
Intensity	Median	GLRLM	Grey level nonuniformity
Intensity	Median absolute deviation	GLRLM	Grey level variance
Intensity	Minimum	GLRLM	High grey level run emphasis
Intensity	Range	GLRLM	Long run high grey level emphasis
Intensity	Robust mean absolute deviation	GLRLM	Long run low grey level emphasis
Intensity	Skewness	GLRLM	Long runs emphasis
Intensity	Uniformity	GLRLM	Low grey level run emphasis
Intensity	Variance	GLRLM	Normalized grey level nonuniformity
Intensity	Root mean square	GLRLM	Normalized run length nonuniformity
GLCM	Angular second moment	GLRLM	Run length nonuniformity
GLCM	Autocorrelation	GLRLM	Run entropy
GLCM	Cluster prominence	GLRLM	Run length variance
GLCM	Contrast	GLRLM	Run percentage
GLCM	Correlation	GLRLM	Short run high grey level emphasis
GLCM	Difference average	GLRLM	Short run low grey level emphasis
GLCM	Difference entropy	GLRLM	Short runs emphasis
GLCM	Difference variance	NGTDM	Busyness
GLCM	Dissimilarity	NGTDM	Coarseness
GLCM	Information correlation 1	NGTDM	Complexity
GLCM	Information correlation 2	NGTDM	Contrast
GLCM	Inverse difference	NGTDM	Texture strength
GLCM	Inverse difference moment		
GLCM	Inverse variance		

Abbreviations. GLCM = Grey Level Co-occurence Matrix, GLRLM = Grey Level Run Length Matrix, NGTDM = Neighbourhood Grey Tone Difference Matrix.

B Regression coefficients other models

Table A.2: Regression coefficients from reference model by *van Rijn-Dekker et al. (2021)* and the regression coefficients of the updated model based on our study population after closed-testing procedure. [89]

Model predictors	**Regression coefficient original**	**Regression coefficient updated**
Intercept	-2.541	-1.894
Pretreatment daytime xerostomia (any)	0.625	0.625
Mean dose to contralateral SCR region*	0.192	0.192
Mean dose to oral cavity	0.018	0.018

* Square root transformation & fixated in model development.

Table A.3: Original regression coefficients from models in article by *van Dijk et al. (2018)* after their external validation. [59]

Model predictors	**Regression coefficient intensity model**	**Regression coefficient texture model**
Intercept	-8.540	-2.673
Pretreatment xerostomia (any)	2.063	2.009
Parotid gland average dose	0.026	0.059
P90	0.054	-
Grey level nonuniformity normalized	-	-0.841

Table A.4: Updated regression coefficients from models by *van Dijk et al. (2018)* after closed-testing procedure on our study data.

Model predictors	**Regression coefficient intensity model**	**Regression coefficient texture model**
Intercept	-4.306	-0.907
Pretreatment xerostomia (any)	1.037	0.992
Parotid gland average dose	0.027	0.029
P90	0.013	-
Grey level nonuniformity normalized	-	-0.415

C Flowchart in- and exclusion

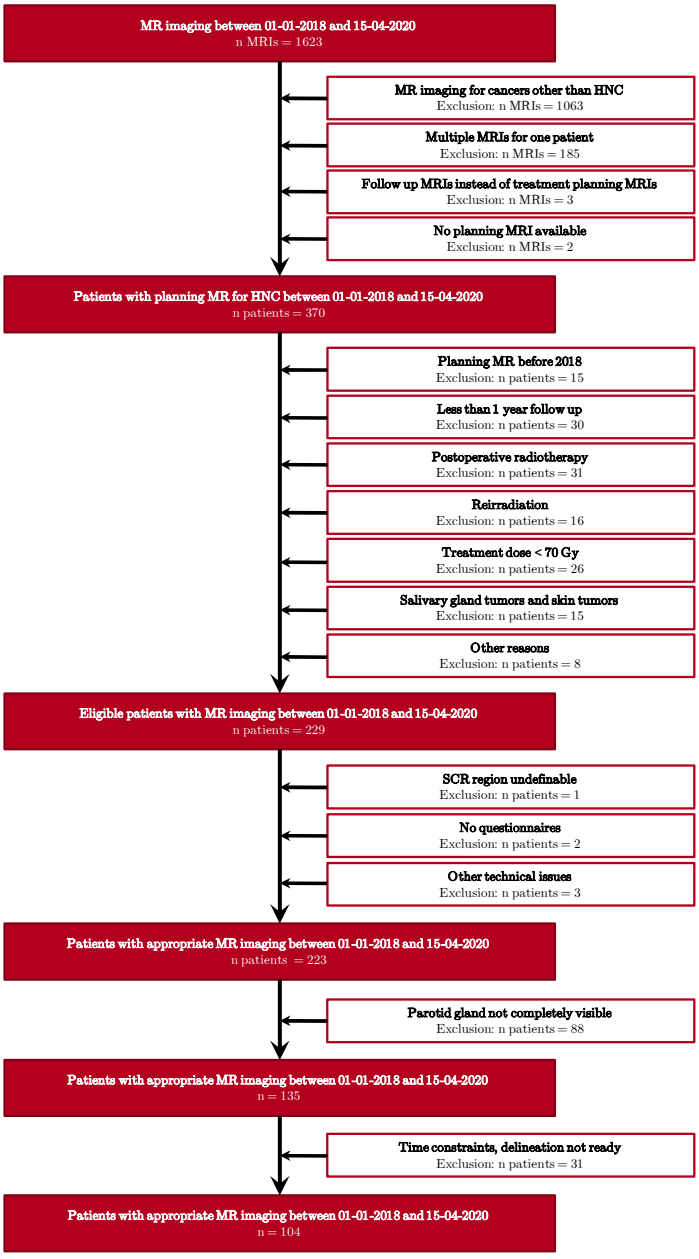


Figure A.1: Flowchart in- and exclusion patients.

D Missing data

Table A.5: Missing data.

	Available	**Missing**
Patient-rated daytime xerostomia baseline	89 (85.6%)	15 (14.4%)
Patient-rated daytime xerostomia M12	66 (63.5%)	38 (36.5%)
Patient-rated general xerostomia baseline	89 (85.6%)	15 (14.4%)
Patient-rated general xerostomia M12	66 (63.5%)	38 (36.5%)

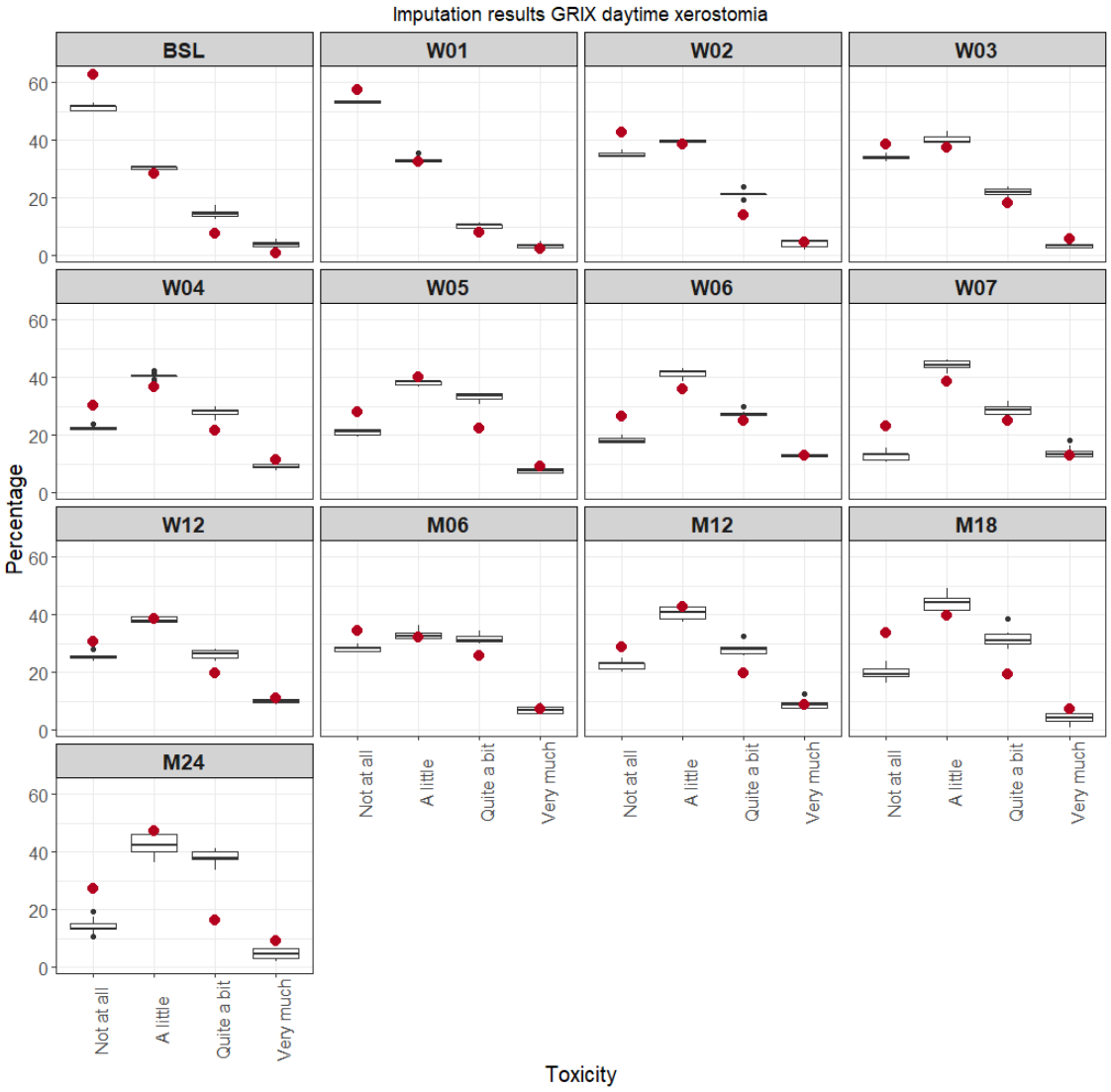
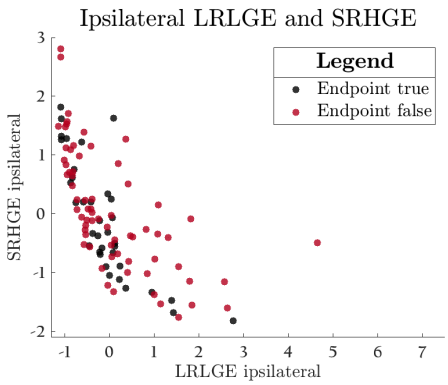
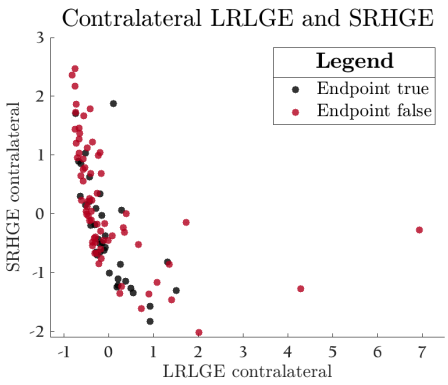


Figure A.2: Imputation results for endpoint (daytime xerostomia at 12 months after treatment) and the other timepoints of this variable. Red dots represent the distribution of all available data prior to imputation. The boxplots represent the distribution of imputated data in 10 imputation sets.
BSL = baseline, prior to treatment. WXX = X$^{\text{th}}$ week since start of treatment. MXX = X$^{\text{th}}$ month after treatment

E Correlation IBMs



(a) Spearman correlation = -0.78



(b) Spearman correlation = -0.8

Figure A.3: Scatterplots of the ipsilateral LRLGE and SRHGE (a.) and the contralateral LRLGE and SRHGE (b.).

F Comparison populations

Table A.6: Comparison of values for predictors and endpoints in our study population and in the study population of the reference model.

Parameter	**Study population reference model n=663**	**Our study population n=104**	**p-value**
Prevalence daytime xerostomia baseline (any)	38.5%[a]	49%	0.07[b]
Prevalence daytime xerostomia M12 (moderate-to-severe)	32.9%[a]	36.8%	0.59[b]
Contralateral SCR mean dose (Gy) Median (IQR)	19.3 (9.2-26.9)	9.09 (5.2-13.4)	<0.001[c]
Ipsilateral SCR mean dose (Gy) Median (IQR)	27.78 (15.19-42.67)	20.62 (13.48-33.22)	0.012[c]
Oral cavity mean dose (Gy) Median (IQR)	45.75 (27.5-55.8)	32.6 (22.8-42.9)	<0.001[c]
Contralateral parotid gland mean dose (Gy) Median (IQR)	25.20 (15.87-32.36)	15.60 (8.87-20.42)	<0.001[c]
Ipsilateral parotid gland mean dose (Gy) Median (IQR)	32.91 (21.99-43.79)	25.46 (19.87-32.96)	0.001[c]
Contralateral submandibular gland mean dose (Gy) Median (IQR)	53.69(46.44-62.43)	42.52 (33.70-56.33)	<0.001[c]
Ipsilateral submandibular gland mean dose (Gy) Median (IQR)	64.80 (50.98-68.45)	64.50 (58.02-68.10)	0.618[c]
Tumour location			
Oral Cavity	5.6 %	8.7 %	-
Oropharynx	39.8 %	71.2 %	-
Nasopharynx	4.4 %	12.5 %	-
Hypopharynx	10.4 %	1.0 %	-
Larynx	39.8 %	2.9 %	-
Miscellaneous	0 %	3.8 %	-

[a] Amount of missing data: daytime xerostomia baseline = 11.4%, daytime xerostomia M12 = 29.2%.
[b] Chi-Squared test.
[c] Kruskall-Wallis test.