

University of Groningen

Predicting salivary gland dysfunction with image biomarkers in head and neck cancer patients

van Dijk, Lisanne Vania

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
van Dijk, L. V. (2018). Predicting salivary gland dysfunction with image biomarkers in head and neck cancer patients. [Groningen]: Rijksuniversiteit Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

**Predicting salivary gland
dysfunction with
image biomarkers in
head and neck cancer patients**

Lisanne Vania van Dijk

Colofon

Predicting salivary gland dysfunction with image biomarkers in head and neck cancer patients

ISBN/EAN: 978-94-028-1225-1

Copyright © 2018 LV van Dijk

All rights reserved. No part of this thesis may be reproduced, stored or transmitted in any way or by any means without the prior permission of the author, or when applicable, of the publishers of the scientific papers.

Layout and design by Matthijs Ariens, persoonlijkproefschrift.nl.

Printed by Ipskamp Printing, proefschriften.net.

Financial support for the publication of this thesis was kindly provided by:

- Mirada Medical B.V.
- Elekta B.V.
- University Medical Center Groningen
- University of Groningen



rijksuniversiteit
 groningen

Predicting salivary gland dysfunction with image biomarkers in head and neck cancer patients

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 28 november 2018 om 16.15 uur

door

Lisanne Vania van Dijk

geboren op 21 november 1988
te 's-Gravenhage

Promotor

Prof. dr. J.A. Langendijk

Copromotores

Dr. R.J.H.M. Steenbakkers

Dr. ir. N.M. Sijtsema

Beoordelingscommissie

Prof. dr. C.R. Leemans

Prof. dr. A. Vissink

Prof. dr. J. Pruim

Content

Chapter 1	General introduction	7
Part 1	Pre-treatment image biomarkers predict late salivary gland dysfunction	
Chapter 2	CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva	17
Chapter 3	¹⁸ F-FDG PET image biomarkers improve prediction of late radiation-induced xerostomia	39
Chapter 4	Parotid gland fat related MR image biomarkers improve prediction of late radiation-induced xerostomia	61
Part 2	Image biomarkers changes after and during radiotherapy predict late xerostomia	
Chapter 5	Geometric image biomarker changes of the parotid gland are associated with late xerostomia	83
Chapter 6	Parotid gland surface area reduction during radiotherapy improves the prediction of late xerostomia	105
Chapter 7	Summary and general discussion	125
Appendices	Nederlandse samenvatting	138
	Dankwoord	142
	Curriculum Vitae	148
	Publications list	150

Chapter 1

General introduction

Introduction thesis

Radiotherapy plays a pivotal role in the treatment of patients with head and neck Cancer (HNC), either as single modality or in combination with systemic treatment and/or surgery [1]. The majority of HNC concerns squamous cell carcinoma and arises from regions in or adjacent to the upper digestive tract. Survival rates have improved in the last decades due to improvement of treatment strategies [2–9]. The introduction of radiotherapy treatment techniques like Intensity Modulated Radiotherapy (IMRT) and Volumetric Modulated Arc Therapy (VMAT) have resulted in more conformal dose distributions and have been successfully combined with systemic agents, including concurrent chemotherapy and cetuximab [2–4]. Moreover, survival rates have improved due to increasing incidences of human papilloma virus (HPV) related HNC, since patients with HPV-positive tumours show a remarkably better overall survival compared to those with a non-HPV related tumours [7–9].

The increased life expectancy of HNC survivors has led to a rising demand for adequate prediction, prevention and understanding of the development of treatment-induced side effects. In addition, more advanced treatment options are becoming available that have great potential to spare normal tissues, such as proton therapy [10,11] and Magnetic Resonance Imaging (MRI) guided radiation [12]. However, these advanced treatment techniques are currently limited available and their benefit varies between patients. In the Netherlands, the model-based approach has been introduced as an evidence-based method to select patients for the most optimal treatment based on differences in the expected toxicity profiles between treatment modalities [10], illustrating how toxicity prediction can contribute to more individualized treatment strategies. Following radiotherapy, the most frequently reported side effects are xerostomia, which is the syndrome of dry mouth, and sticky saliva, due to changes of saliva composition, and swallowing dysfunction (dysphagia) [1,13,14]. These toxicities normally become clinically apparent during radiotherapy (35 radiation fractions in 6 or 7 weeks) [13]. These side effects may persist for weeks or months after treatment, while in some patients no recovery is observed leading to burdensome complaints for the rest of their lives. [15]. Figure 1 depicts the incidences of patient-reported moderate-to-severe xerostomia during and after completion of treatment from a combined cohort of patients included in this thesis, and are obtained from our department's Standard Follow-up Program. Especially, these

late side effects are very disabling for patients, and have a major impact on the quality of life of HNC patients [14].

This thesis focusses on late side effects related to salivary gland dysfunction, xerostomia and sticky saliva. These side effects may also lead to altered taste, dental infection, swallowing, and speech problems [15]. Multiple studies have shown that the radiation dose administered to the parotid glands, which are the major salivary glands (figure 2), is associated with the development of late xerostomia [16–21], while submandibular gland doses were found to be related to the development of sticky saliva after treatment [16,17].

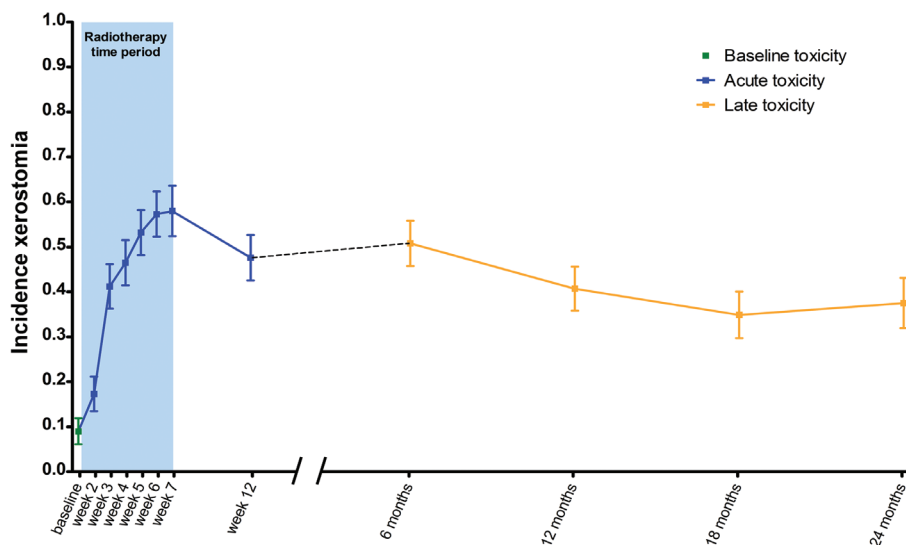


Figure 1 Example of toxicity development. Moderate-to-severe xerostomia incidences before, weekly during and 6 weeks, 6, 12, 16 and 24 months after radiotherapy of a sample size of 396 HNC patients from a combined cohort of patients included in this thesis.

To predict side effects, Normal Tissue Complication Probability (NTCP) models are used. NTCP-models are prediction models that describe the relationship between 3D-dose distributions and the risk on radiation-induced side effect. Several studies have presented univariable NTCP models based on mean dose to the parotid glands that predict the reduction of salivary flow rates below 25% [19–21]. Houweling et al. compared several model types (e.g. Lyman-Kutcher-Burman, mean dose exponential and dose-threshold model) and showed that the logistic regression model based on mean dose to both parotid glands performed best predicting salivary flow reduction [18]. Reduced salivary flow, however does not necessarily translate in altered patient-reported outcomes (PRO) [22]. Beetz



et al. were the first to develop a multivariable NTCP model predicting late patient-rated xerostomia after radiotherapy [17]. The predictors were mean dose to the contralateral parotid gland and baseline xerostomia scores. In addition, sticky saliva prediction was based on mean dose to the submandibular and sublingual glands and soft palate [17].

However, substantial unexplained variability in predicting xerostomia and sticky saliva remains for these conventional NTCP models that are based on dose-volume parameters and baseline toxicity scores. In other words, patients receiving similar radiation doses and with similar baseline complaints can react very differently to treatment. Optimisation of the performance of NTCP models is a necessary next step to further support more personalised treatment approaches.

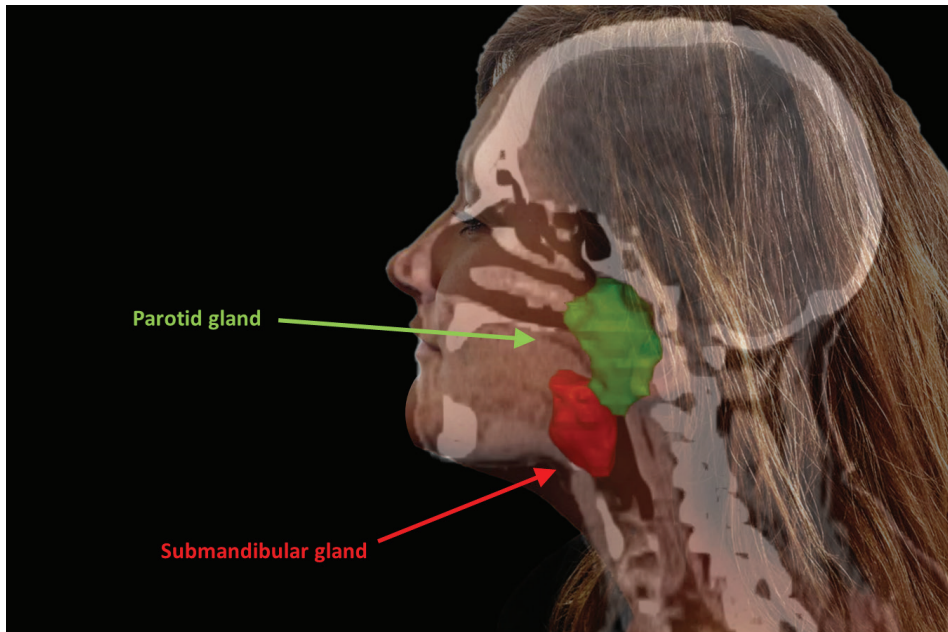


Figure 2 Anatomical representation of the parotid and submandibular gland.

In this thesis, we tested the hypothesis that the prediction of radiation-induced salivary gland toxicities can be improved by adding patient-specific information extracted from 3-dimensional images, such as Computed Tomography (CT), Positron Emission Tomography (PET) or Magnetic Resonance Imaging (MRI). These images are routinely acquired for delineation (i.e. tumour and organs at

risk segmentation) and treatment planning purposes (Figure 3), yet containing additional unused information of patient's anatomy and physiology.

Radiomics refers to the process of converting medical images into high-dimensional minable data [23]. Patient-specific tissue characteristics are quantified in so-called image biomarkers (IBMs) or features. They represent intensity, texture and geometric properties of tissue from a specific volume of interest. Aerts et al. showed that CT image biomarkers describing the density, heterogeneity and shape of the tumour, could predict overall survival of non-small cell lung cancer patients and validated this in both an independent HNC and lung cancer patient cohort [24]. Subsequently, several studies have shown that tumour image biomarkers can contribute to the prediction of overall, disease-free and progression-free survival in HNC patients [24–28]. However, so far, the role of these image biomarkers extracted from normal tissues to predict radiation-induced toxicities is less explored, while these are imperative in supporting treatment decisions [10].

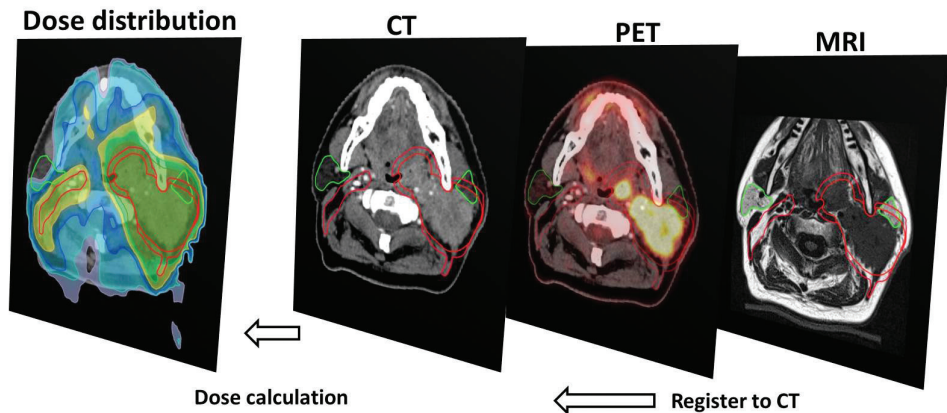


Figure 3 Currently standard in radiotherapy, CT is used for the dose distribution calculation and for the delineation of the target regions and organs at risk. ^{18}F -FDG PET and MR images are often registered to CT to provide additional information, such as metabolic activity and superior soft tissue contrast, for tumour delineation. Delineations of clinical target volumes (red) and the parotid glands (green) are depicted.



Outline of the thesis

The aim of this thesis was to improve the prediction of radiation-induced salivary gland toxicities in HNC patients with normal tissues image biomarkers, by adding them to conventional NTCP models that are based on dose-volume parameters and baseline complaints only.

The first part of this thesis (chapter 2-4) focuses on improving the prediction of late toxicities with image biomarkers that are extracted from pre-treatment images. Optimized pre-treatment prediction is necessary to identify patients that are most at risk of developing persistent salivary dysfunction and thus may be good candidates for more advanced treatment techniques, such as proton therapy and MRI-guided radiation [11,12], which could further support more effective personalized treatment approaches.

The second part of this thesis (chapter 5-6) focusses on identifying parotid gland changes observed during and early after treatment, which were quantified in Δ image biomarkers and associated with late xerostomia. Quantification of normal tissue changes in an early stage that are associated with permanent damage could identify patients that will not recover and could potentially guide treatment adaptation to prevent late toxicities as much as possible.

Chapter 2 aims to improve the prediction of late xerostomia and sticky saliva by investigating image biomarkers of parotid and submandibular glands in pre-treatment CT images. CT is the most apparent modality to investigate first, since CT images are always acquired for radiotherapy treatment planning and give a stable representation of the tissue density.

Chapter 3 investigates the improvement of toxicity prediction with the metabolic activity of the parotid gland by extracting image biomarkers from pre-treatment ^{18}F -FDG PET images. This image modality gives a spatial distribution of glucose (FDG) labelled with a radioactive marker (^{18}F) in patients, which relates to the local metabolic activity in the tissue.

Chapter 4 tests the hypothesis resulting from chapter 2 and 3 that fat in the parotid gland is a xerostomia risk factor by investigating whether image biomarkers, extracted from pre-treatment T1-weighted MR images, are associated with the development of late xerostomia. Although MRI is a complex image modality, it is the most preferred modality to support the hypothesis due to its excellent soft tissue contrast.

Chapter 5 investigates the relation of parotid gland dose with parotid gland changes, quantified by Δ image biomarkers, before and 6 weeks after radiotherapy, together with the association of these Δ image biomarkers to late xerostomia.

Chapter 6 identifies predictive Δ image biomarkers during treatment that can be used to identify patients at risk for late xerostomia, early in-treatment.

The findings described in this thesis are summarized and discussed in **Chapter 7**.



References

- 1 Vissink A, Mitchell JB, Baum BJ, Limesand KH, Jensen SB, Fox PC, et al. Clinical management of salivary gland hypofunction and xerostomia in head-and-neck cancer patients: Successes and barriers. *Int J Radiat Oncol Biol Phys* 2010;78:983–91.
- 2 Beadle BM, Liao K-P, Elting LS, Buchholz T a., Ang KK, Garden AS, et al. Improved survival using intensity-modulated radiation therapy in head and neck cancers: A SEER-Medicare analysis. *Cancer* 2014;120:702–10.
- 3 Bonner JA, Harari PM, Giralt J, Cohen RB, Jones CU, Sur RK, 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. *Lancet Oncol* 2010;11:21–8.
- 4 Pignon JP, Maître A le, Maillard E, Bourhis J. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): An update on 93 randomised trials and 17,346 patients. *Radiother Oncol* 2009;92:4–14.
- 5 Pulte D, Brenner H. Changes in survival in head and neck cancers in the late 20th and early 21st century: a period analysis. *Oncologist* 2010;15:994–1001.
- 6 Karim-Kos HE, de Vries E, Soerjomataram I, Lemmens V, Siesling S, Coebergh JWW. Recent trends of cancer in Europe: A combined approach of incidence, survival and mortality for 17 cancer sites since the 1990s. *Eur J Cancer* 2008;44:1345–89.
- 7 Marur S, D’Souza G, Westra WH, Forastiere AA. HPV-associated head and neck cancer: A virus-related cancer epidemic. *Lancet Oncol* 2010;11:781–9.
- 8 Braakhuis BJM, Visser O, René Leemans C. Oral and oropharyngeal cancer in The Netherlands between 1989 and 2006: Increasing incidence, but not in young adults. *Oral Oncol* 2009;45:e85–9.
- 9 Benson E, Li R, Eisele D, Fakhry C. The clinical impact of HPV tumor status upon head and neck squamous cell carcinomas. *Oral Oncol* 2014;50:565–74.
- 10 Langendijk JA, Lambin P, De Ruyscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. *Radiother Oncol* 2013;107:267–73.
- 11 Lomax A. Intensity modulation methods for proton radiotherapy. *Phys Med Biol* 1999;44:185–205.
- 12 Lagendijk JJW, Raaymakers BW, Raaijmakers AJE, Overweg J, Brown KJ, Kerkhof EM, et al. MRI/linac integration. *Radiother Oncol* 2008;86:25–9.
- 13 Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide S a, Clark C, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. *Lancet Oncol* 2011;12:127–36.
- 14 Langendijk JA, Doornaert P, Verdonck-de Leeuw IM, Leemans CR, Aaronson NK, Slotman BJ. Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. *J Clin Oncol* 2008;26:3770–6.

- 15 Wijers OB, Levendag PC, Braaksma MM, Boonzaaijer M, Visch LL, Schmitz PI. Patients with head and neck cancer cured by radiation therapy: a survey of the dry mouth syndrome in long-term survivors. *Head Neck* 2002;24:737–47.
- 16 Jellema AP, Doornaert P, Slotman BJ, Leemans CR, Langendijk J a. Does radiation dose to the salivary glands and oral cavity predict patient-rated xerostomia and sticky saliva in head and neck cancer patients treated with curative radiotherapy? *Radiother Oncol* 2005;77:164–71.
- 17 Beetz I, Schilstra C, Van Der Schaaf A, Van Den Heuvel ER, Doornaert P, Van Luijk P, 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. *Radiother Oncol* 2012;105:101–6.
- 18 Houweling AC, Philippens MEP, Dijkema T, Roesink JM, Terhaard CHJ, Schilstra C, et al. A comparison of dose-response models for the parotid gland in a large group of head-and-neck cancer patients. *Int J Radiat Oncol Biol Phys* 2010;76:1259–65.
- 19 Eisbruch A, Kim HM, Terrell JE, Marsh LH, Dawson LA, Ship JA. Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2001;50:695–704.
- 20 Roesink JM, Moerland MA, Battermann JJ, Hordijk GJ, Terhaard CHJ. Quantitative dose-volume response analysis of changes in parotid gland function after radiotherapy in the head-and-neck region. *Int J Radiat Oncol* 2001;51:938–46.
- 21 Dijkema T, Terhaard CHJ, Roesink JM, Braam PM, van Gils CH, Moerland M a, et al. Large cohort dose-volume response analysis of parotid gland function after radiotherapy: intensity-modulated versus conventional radiotherapy. *Int J Radiat Oncol Biol Phys* 2008;72:1101–9.
- 22 Saleh J, Figueiredo MAZ, Cherubini K, Salum FG. Salivary hypofunction: An update on aetiology, diagnosis and therapeutics. *Arch Oral Biol* 2014;60:242–55.
- 23 Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2015;278:151169.
- 24 Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5.
- 25 Abgral R, Keromnes N, Robin P, Le Roux P-Y, Bourhis D, Palard X, et al. Prognostic value of volumetric parameters measured by (18)F-FDG PET/CT in patients with head and neck squamous cell carcinoma. *Eur J Nucl Med Mol Imaging* 2014;41:659–67.
- 26 Koyasu S, Nakamoto Y, Kikuchi M, Suzuki K, Hayashida K, Itoh K, et al. Prognostic value of pretreatment 18F-FDG PET/CT parameters including visual evaluation in patients with head and neck squamous cell carcinoma. *AJR Am J Roentgenol* 2014;202:851–8.
- 27 Alluri KC, Tahari AK, Wahl RL, Koch W, Chung CH, Subramaniam RM. Prognostic value of FDG PET metabolic tumor volume in human papillomavirus-positive stage III and IV oropharyngeal squamous cell carcinoma. *AJR Am J Roentgenol* 2014;203:897–903.
- 28 Zhai T-T, van Dijk L V, Huang B-T, Lin Z-X, Ribeiro CO, Brouwer CL, et al. Improving the prediction of overall survival for head and neck cancer patients using image biomarkers in combination with clinical parameters. *Radiother Oncol* 2017:256–62.

Chapter 2

CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva

Published in: **Radiotherapy and Oncology** 2016 July; 122:185-91

van Dijk LV, Brouwer CL, van der Schaaf A, Burgerhof JGM, Beukinga RJ, Langendijk JA, Sijtsema NM, Steenbakkens RJHM.

Online Supplemental Materials : <https://doi.org/10.1016/j.radonc.2016.07.007>

Abstract

Background and purpose

Current models for the prediction of late patient-rated moderate-to-severe xerostomia (XER_{12m}) and sticky saliva ($STIC_{12m}$) after radiotherapy are based on dose-volume parameters and baseline xerostomia (XER_{base}) or sticky saliva ($STIC_{base}$) scores. The purpose is to improve prediction of XER_{12m} and $STIC_{12m}$ with patient-specific characteristics, based on CT image biomarkers (IBMs).

Materials and Methods

Planning CT-scans and patient-rated outcome measures were prospectively collected for 249 head and neck cancer patients treated with definitive radiotherapy with or without systemic treatment. The potential IBMs represent geometric, CT intensity and textural characteristics of the parotid and submandibular glands. Lasso regularisation was used to create multivariable logistic regression models, which were internally validated by bootstrapping.

Results

The prediction of XER_{12m} could be improved significantly by adding the IBM "Short Run Emphasis" (SRE), which quantifies heterogeneity of parotid tissue, to a model with mean contra-lateral parotid gland dose and XER_{base} . For $STIC_{12m}$, the IBM maximum CT intensity of the submandibular gland was selected in addition to $STIC_{base}$ and mean dose to submandibular glands.

Conclusion

Prediction of XER_{12m} and $STIC_{12m}$ was improved by including IBMs representing heterogeneity and density of the salivary glands, respectively. These IBMs could guide additional research to the patient-specific response of healthy tissue to radiation dose.

Introduction

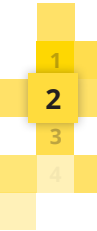
The survival of head and neck cancer (HNC) patients has improved remarkably in the last decade with the addition of systemic agents, including concurrent chemotherapy and cetuximab [1,2]. However, these treatment strategies have significantly increased acute and late toxicity [3]. Consequently, reducing treatment-induced side effects has become increasingly important. Despite the clinical introduction of more advanced radiation techniques, side effects related to hyposalivation, such as xerostomia and sticky saliva, are still frequently reported following radiotherapy (RT) for HNC. Accurate prediction of these side effects is important in order to individually tailor treatments to patients.

To predict moderate-to-severe xerostomia and sticky saliva, Normal Tissue Complication Probability (NTCP) models have been developed [4,5]. Current models are based on a combination of dose-volume parameters of salivary glands and baseline risk factors. However, these models cannot completely explain the variation in development of xerostomia between individuals. Therefore, identification of additional factors is needed to explain the patient-specific response to dose, and subsequently to optimize NTCP models.

In current clinical practice, three-dimensional anatomic information is acquired with planning CT scans for all patients receiving RT. These scans are used to delineate the target and organs at risk, and to calculate the dose distribution of the planned treatment. These scans yield reproducible information about patient-specific anatomy and tissue composition, and could therefore contribute to the understanding and prediction of the development of side effects in HNC patients.

Information about the structure, shape and composition of organs at risk from the CT can be quantified with image features. Features that correlate with treatment outcome or complications can be used as so called image biomarkers (IBMs). Extracted from CT data of the parotid (PG) and submandibular glands (SG), the different image features represent their CT intensity as well as geometric and textural characteristics.

Aerts et al. [6] investigated the relationship between CT IBMs of head and neck tumours and survival. Furthermore, the relationship between geometric changes of organs at risk after RT, and radiation induced complications, has been described in several studies [7–10]. Scalco et al. [11] investigated change after RT for a selected set of textural parameters. However, there are no studies so far



that report on the relationship between IBMs of organs at risk before treatment and the risk of complications.

The aim of this study, therefore, was to investigate the prediction of xerostomia and sticky saliva, as assessed at 12 months after radiotherapy. The objective was to improve predictions by the addition of IBMs of the parotid and submandibular glands, determined from the planning CT-scans, to models that contain clinical and dosimetric information.

Method

Patient demographics and treatment

The study population of HNC patients was treated with definitive radiotherapy either in combination or not with concurrent chemotherapy or cetuximab, between July 2007 and August 2014. Patients with tumours in the salivary glands, those with excised parotid or submandibular glands and/or patients that underwent surgery in the head and neck area were excluded from this study. Furthermore, patients with metal streaking artifacts in the CT were excluded, due to the influence of CT intensity values that do not correspond to tissue densities. Moreover, patients without follow-up data 12 months after RT were also excluded. Patient characteristics are depicted in Table 1.

For each patient, a planning CT (Somatom Sensation Open, Siemens, Forchheim, Germany, voxel size: $0.94 \times 0.94 \times 2.0 \text{ mm}^3$; 100-140 kV) with contrast enhancement was acquired. This CT was used for contouring and RT planning. The parotid and submandibular glands were delineated according to guidelines as described by Brouwer et al. [12].

Most patients were treated with standard parotid sparing IMRT (ST-IMRT) or swallowing sparing IMRT (SW-IMRT) [13,14]. All IMRT and VMAT treatments included a simultaneous integrated boost (SIB) and attempted to spare the parotid glands and/or the swallowing structures without compromising the dose to the target volumes [15]. The tumour and, if present, pathological lymph node target volumes, received a total dose of 70 Gy (2 Gy per fraction). Most patients received an elective radiation dose of 54.25 Gy (1.55 Gy per fraction) on the lymph node levels that were delineated as described by Gregoire et al. [16]. Radiation protocols were similar to those described by Christianen et al. [17].

Table 1 Patient characteristics

Characteristics	N=249	%
Sex		
Female	61	24
Male	188	76
Age		
18 - 65 years	133	53
> 65 years	116	47
Tumour site		
Oropharynx	74	30
Nasopharynx	14	6
Hypopharynx	31	12
Larynx	118	47
Oral cavity	11	4
Unknown primary	1	0
Tumour classification		
T0	3	1
T1	27	11
T2	81	33
T3	77	31
T4	61	24
Node classification		
N0	115	46
N1	23	9
N2abc	104	42
N3	7	3
Systemic treatment		
yes	100	40
no	149	60
Treatment technique		
3D-CRT	23	9
ST-IMRT	92	37
SW-IMRT	124	50
SW-VMAT	10	4
Bi-lateral		
yes	203	82
no	46	18

Abbreviations: CRT: Conformal Radiation Therapy; IMRT: Intensity-Modulated Radiation Therapy; ST-IMRT: standard parotid sparing IMRT; SW-IMRT: swallowing sparing IMRT; SW-VMAT: swallowing sparing Volumetric Arc Therapy



Endpoints

The EORTC QLQ-H&N35 questionnaire was used to evaluate patient-rated xerostomia and sticky saliva before and after RT. This questionnaire is part of a standard follow-up programme (SFP), as described in previous reports [4,18,19], and uses a 4-point Likert scale that describes the condition as 'none', 'a bit', 'quite a bit' and 'a lot'. All patients included were subjected to the SFP programme, where toxicity and quality of life were evaluated prospectively on a routine basis; before, during and after treatment.

The endpoints of this study are moderate-to-severe xerostomia (XER_{12m}) and sticky saliva (STIC_{12m}) 12 month after RT. This corresponds to the 2 highest scores on the 4-point Likert scale.

Potential CT image biomarkers, dose and clinical parameters

Dose and clinical parameters

The planning CT, dose distribution and delineated structures were analysed in Matlab (version R2014a). Both the mean dose to the contra- and bi-lateral parotid and submandibular glands were determined, since previous studies have shown that those were the most important parameters in the prediction of patient-rated xerostomia and sticky saliva at 6 and 12 months after RT [4,5,20].

Furthermore, different patient characteristics (age, sex, WHO-stage, weight, length and Body Mass Index), tumour characteristics (TNM stage, tumour location) and treatment characteristics (treatment technique and the use of systemic treatment) were also included. In addition, the patient-rated xerostomia and sticky saliva at baseline were taken into account.

CT intensity and geometric image biomarkers

Patient-specific characteristics of the parotid and submandibular glands were quantified by extracting potential CT IBMs, representing geometric, CT-intensity and pattern characteristics. In figure 1, extraction of different types of IBMs is explained schematically. The in-house developed software that was used to extract the IBMs was based on commonly used formulas (supplementary data 1 and 2) and implemented in Matlab (version R2014a). The CT intensity IBMs (number = 24) were derived from the CT intensity information of the delineated volumes of interest. Examples of these features are mean, variance, minimum,

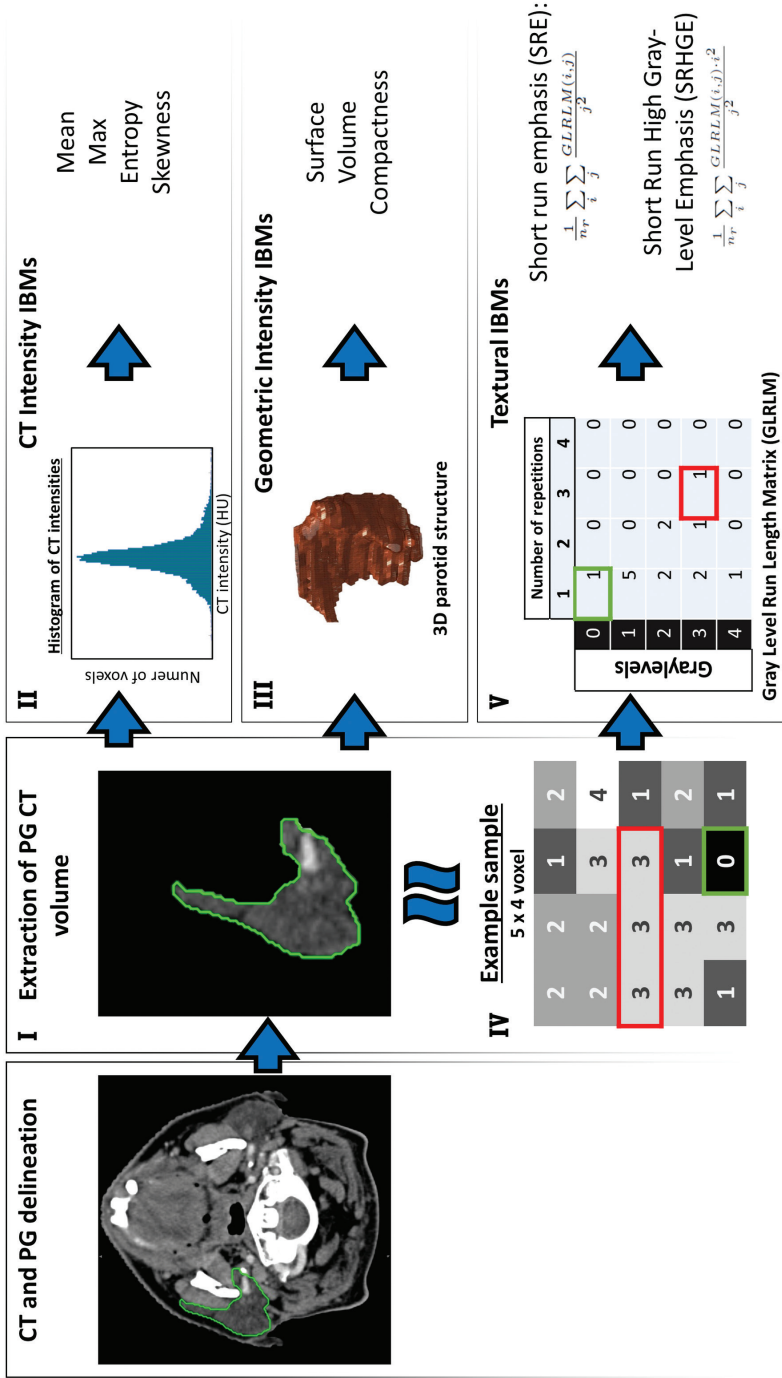


Figure 1 Examples of the Image Biomarker (IBM) extraction process. The delineated gland of interest is extracted from the CT image (I). CT intensity IBMs are obtained from all voxels inside the contour (II). Geometric IBMs are derived from the delineation of the gland directly (III). A small sample of the CT where voxel intensity values are binned (IV). In this example, a GLRLM matrix is constructed from this CT data by quantifying the number of repetitions of gray intensities from left to right (V).



maximum, quantiles, energy and skewness of CT intensity. The geometric IBMs (number = 20), such as volume, sphericity, compactness and major and minor axis length, were directly derived from the delineated structures.

Textural image biomarkers

More complex CT IBMs are defined to describe the heterogeneity of tissue. These textural IBMs (number = 86) were derived from the gray level co-occurrence matrix (GLCM) [21], gray level run-length matrix (GLRLM) [22] and gray level size-zone matrix (GLSZM) [23]. To extract this, the CT intensities were binned from -200 to 200 Hounsfield Units (HU) with an interval of 25 HU. All textural features were normalized by subtracting the IBM values from their mean and dividing by the standard deviation. For more information on textural IBM extraction, refer to supplementary data 2 and Aerts et al. [6]. Ultimately, all potential CT IBMs and clinical and dosimetric parameters together resulted in 142 variables.

Pre-selection of variables and univariable analysis

A large number of potential variables can increase the risk of false positives, overfitting the model and of multicollinearity [24,25]. In this study, a method for pre-selecting variables was applied to reduce the probability of these adverse effects. First, the (Pearson) correlation was determined between all combinations of variables. If a correlation larger than 0.80 was observed, then the variable with the lowest univariable correlation with the endpoint was omitted. After pre-selection, univariable analysis of the pre-selected variables was performed.

Multivariable analysis and model performance

Lasso regularisation was used to create two multivariable logistic regression models to predict moderate-to-severe XER_{12m} and STIC_{12m}. All pre-selected variables were introduced to the modelling process. By increasing the penalisation term lambda, the regularisation shrinks the coefficients of the variables and thereby excludes variables by reducing them to zero. To robustly determine the optimal lambda that results in a model that best fits the observed data, 10-fold cross validation was used [26]. This was repeated 100 times, as these folds are randomly picked [26].

General lasso tends to select models with too many variables [27]. Therefore, the 75th quartile (not the average) of the 100 obtained optimal lambdas was used to select the variables [28]. Subsequently, the variables selected by lasso

were again fitted to the data with logistic regression and internally validated through bootstrapping. This validation corrects for optimism by shrinking the model (slope and intercept) and the model performance accordingly [25,29]. Reference models without IBMs were created and the contribution of IBMs to the models was tested with the log-likelihood-ratio test. The model's performance was quantified in terms of discrimination with the Area Under the Curve of the ROC curve (AUC), the Nagelkerke R^2 and the discrimination slope. The Hosmer-Lemeshow test evaluated the calibration. Variance Inflation Factor (VIF) was used to evaluate the correlation of a variable with all others in the model [30]. The R-packages Lasso and Elastic-Net Regularized Generalized Linear Models (version 2.0-2) [26] and Regression Modeling Strategies (version 4.3-1) [31] were used.

Impact of variation in delineation

Delineation of organs at risk in the head and neck region by different observers may be subject to inter-observer variability [32], which could result in a variation in IBM values. To evaluate this, four additional delineations per gland per patient were created by eroding the original delineation by magnitudes corresponding to the variations in delineation reported by Brouwer et al. [32]. The IBM stability was evaluated combining the intra-class correlation of the IBM values of the original and created delineations. An IBM with an intra-class correlation higher than 0.70 was considered relatively stable (1.0 indicates identical observations). For more details, refer to Supplementary data 3.

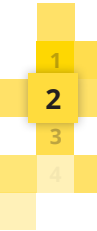
Results

Patients

After exclusion of patients with metal artefacts in the CT-scans, 424 of the 629 patients (67%) were included. Of the remaining patients, 249 (39%) completed the EORTC QLQ-HN35 at 12 months after treatment and were included in the analysis. Moderate-to-severe xerostomia was reported in 40% (100) and sticky saliva in 25% (63) of these patients.

Pre-selecting variables and univariable analysis

After testing of inter-variable correlation (Pearson), a selection of 26 of 142 variables for XER_{12m} and 24 of 142 variables for $STIC_{12m}$ were pre-selected. Univariable analysis of the pre-selected variables showed that 8 and 6 variables



were significantly correlated to XER_{12m} and $STIC_{12m}$, respectively (p -value < 0.05) (Table 2). However, all pre-selected variables were used in the lasso regularisation process. These pre-selected variables are listed in the supplementary data 4.

Table 2 (part 1) Univariable analysis after pre-selection of parotid gland related variables for xerostomia

Xerostomia at 12 months after RT				
Name	Type	β	p-value	OR (95% CI)
Mean dose contra (PG)	DVH	0.06	<0.001	1.06 (1.04-1.09)
Baseline xerostomia	Clinical	0.80	<0.001	2.22 (1.49-3.30)
Short Run Emphasis	GLRLM	0.44	0.002	1.55 (1.18-2.03)
97.5 percentile	Intensity	0.39	0.004	1.47 (1.13-1.92)
Long Run Emphasis	GLRLM	-0.50	0.014	0.61 (0.41-0.90)
SRHGE	GLRLM	-17.14	0.014	0.00 (0.00-0.03)
Tumour stage	Clinical	0.26	0.039	1.29 (1.01-1.65)
Bounding box volume	Geometric	-0.27	0.046	0.76 (0.59-0.99)

Abbreviations: PG: parotid gland; OR: odds ratio; CI: confidence interval; SRHGE: Short Run High Gray Emphasis

Table 2 (part 2) Univariable analysis after pre-selection of submandibular gland related variables for sticky saliva.

Sticky saliva at 12 months after RT				
Name	Type	β	p-value	OR (95% CI)
Baseline sticky saliva	Clinical	0.99	<0.001	2.70 (1.81-4.03)
Mean dose (SGs)	DVH	0.04	<0.001	1.04 (1.02-1.06)
Maximum	Intensity	0.01	0.001	1.01 (1.00-1.01)
97.5 percentile	Intensity	0.02	0.008	1.02 (1.00-1.03)
Squared homogeneity	GLCM	-0.33	0.027	0.72 (0.54-0.96)
SRHGE	GLRLM	-0.58	0.032	0.56 (0.33-0.95)

Abbreviations: SGs: submandibular glands; OR: odds ratio; CI: confidence interval; SRHGE: Short Run High Gray Emphasis

Multivariable analysis and model performance

For Xer_{12m} , the variables selected by the lasso modelling process were mean dose to the contra-lateral parotid gland, baseline xerostomia and the image biomarker “Short Run Emphasis” (SRE). The SRE significantly improved the model in terms of overall and discrimination performance (Likelihood Ratio test: $p=0.01$). The

AUC increased from 0.75 (0.69-0.81) to 0.77 (0.71-0.82) and the discrimination slope from 0.19 to 0.21.

For STIC_{12m}, the mean dose of both submandibular glands, baseline sticky saliva, the maximum CT intensity and Short Run High Gray Emphasis (SRHGE) were selected. The maximum CT intensity added significantly to the model (Likelihood Ratio test, $p=0.005$). However, the SRHGE did not improve the model performance significantly (log-likelihood-test, $p=0.12$) and had negligible effect on the AUC. Therefore, the variable SRHGE was discarded from further analysis and only the maximum intensity was used. Adding this IBM to the mean dose and baseline sticky saliva based model improved the discrimination slope of the model (from 0.15 to 0.18) and the AUC (from 0.74 (0.67-0.80) to 0.77 (0.71-0.83), from 0.73 to 0.76 when tested in bootstrapped data). Resulting (corrected) coefficients and performance measures of the models are depicted in tables 3 and 4, respectively. For the formulas of the final models refer to supplementary data 5.

The Hosmer–Lemeshow test showed that calibration was satisfactory for all models (table 4), indicating a good agreement between the predicted and observed patient outcomes. Additionally, the variance inflation factor (VIF) of all selected variables was < 1.03 , indicating low correlation.

Impact of variation in delineation

For all 249 patients, 4 extra delineations were created of both the contra-lateral parotid and submandibular gland. IBMs were extracted from all delineations. Their robustness was determined with the intra-class correlation (>0.70). For the parotid gland, 92 of all 130 IBMs (71%) were robust. For the submandibular gland, 73 IBMs (56%) were robust. The intra-class correlation of the SRE (IBM in final model Xer12m) was 0.85 (95% CI; 0.82-0.87), indicating that this IBM was relatively robust for contour variations. The maximum intensity of the submandibular gland (IBM in final model STIC12m) was more sensitive for contour variation with an ICC of 0.70 (95% CI; 0.66-0.75).

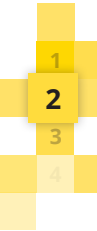


Table 3 Estimated coefficients (uncorrected and corrected for optimism) of NTCP models with and without IBMs

	Model without IBM (Model 1 & 3)				Model with IBM (Model 2 & 4)			
	β	OR (95% CI)	p-value	Average (SD)	β	OR (95% CI)	p-value	Average (SD)
Xerostomia								
Intercept	-3.30	-3.26	-3.31	-3.18				
Contra dose (PG)	0.062	1.06 (1.04-1.09)	<0.001	0.061	0.059	1.06 (1.04-1.09)	<0.001	25.54 (14.38)
XER baseline	0.80	2.23 (1.46-3.41)	<0.001	0.81	0.77	2.24 (1.45-3.45)	<0.001	1.51 (0.68)
SRE GLRLM (PG)	-	-	-	0.40	0.38	1.49 (1.09-2.02)	0.011	0.77*(0.028)
Sticky Saliva								
Intercept	-4.29	-4.24	-4.49	-4.29				
Mean dose (SGs)	0.034	0.033	0.035	0.033	0.033	1.04 (1.01-1.06)	0.005	51.09 (21.34)
STIC baseline	0.86	0.85	0.91	0.86	0.86	2.47 (1.63-3.77)	<0.001	1.47 (0.72)
Max HU (SG)	-	-	0.0077	0.0073	0.0073	1.01 (1.00-1.01)	0.002	177.31 (65.94)

*based on unnormalised values. Abbreviations: Max: maximum; XER: xerostomia ; STIC: sticky saliva; PG: parotid gland; SGs: submandibular glands; SRE: short run emphasis; OR: odds ratio; IBM: image biomarkers; CI: confidence interval

Table 4 Performance of NTCP models with and without IBMs

	Xerostomia				Sticky saliva	
	Model without IBM Model 1	Model with IBM Model 2	Model without IBM Model 3	Model with IBM Model 4	Model without IBM Model 3	Model with IBM Model 4
Overall	-2LLH 283 R ² 0.26	276 0.29	244 0.21	234 0.26		
Discrimination	AUC 0.75 (0.69-0.81) DS 0.19	0.77 (0.71-0.82) 0.21	0.74 (0.67-0.80) 0.15	0.77 (0.71-0.83) 0.18		
Calibration	HL X ² 8.31 HL p-value 0.40	10.98 0.20	9.51 0.30	5.87 0.66		
Validation	AUC boot 0.74 R ² boot 0.25	0.76 0.27	0.73 0.20	0.76 0.24		

Abbreviations: -2LL: -2 log-likelihood; R²: Nagelkerke R²; AUC: Area Under the Curve of the ROC; DS: Discrimination slope; HL: Hosmer-Lemeshow; Boot: corrected for optimism with bootstrapping; IBM: Image Biomarker



Discussion

The results of this study showed that prediction of XER_{12m} and $STIC_{12m}$ could be significantly improved by adding the IBMs short run emphasis (SRE) of the parotid gland and maximum CT intensity of the submandibular gland to the reference models based on dose-volume parameters and baseline factors. The improvements of both models with IBMs persisted when internally validated with both lasso regularisation and bootstrapping. These models with IBMs are a first step to understanding the patient-specific response of healthy tissue to dose. This could contribute to a better prediction of side effects and selection of patients, based on these predictions for advanced treatment techniques, as proposed by Langendijk et al. with the model-based approach to select patients for proton therapy [33].

Short Run Emphasis (SRE) and xerostomia

The SRE obtained from the GLRLM matrix, was associated with the development of XER_{12m} . This IBM is related to the occurrence of short lengths of similar CT intensity value repetitions within the contour. High SRE values indicate heterogeneous parotid tissue or, in other words, that the parotid gland parenchyma is irregular in these patients. Visual investigation of the parotid glands of several patients with high and low SRE suggested that this irregularity resulted from fat saturation of parotid glands (figure 2A-D). The relationship between fat saturation and impaired parotid function has been shown by Izumi et al. [34] for patients with xerostomia related diseases: Sjögren's syndrome and hyperlipidemia. Apparently, the ratio between fatty tissue and functional parotid parenchyma tissue is related to parotid function. Our results suggest that patients with a larger ratio of fat to parotid parenchyma tissue in the parotid glands have a larger risk of developing radiation-induced xerostomia. Our results suggest that patient-specific risk of developing radiation-induced xerostomia can be quantified by IBMs, a first step to explaining the patient-specific response in developing xerostomia to dose. However, CT is not the most optimal image modality to differentiate fat and gland parenchyma. Since MRI is superior in differentiating fat and gland tissue, evaluating parotid glands prior to treatment using MRI images could provide better information for predicting XER_{12m} [35]. Some studies have found a relationship between the initial size of the parotid gland and function prior to [34] and after RT [10,36]. We could not reproduce this

in our population. Only a univariable significant association was found between the volume of the surrounding bounding box of the parotid gland and XER_{12m} .

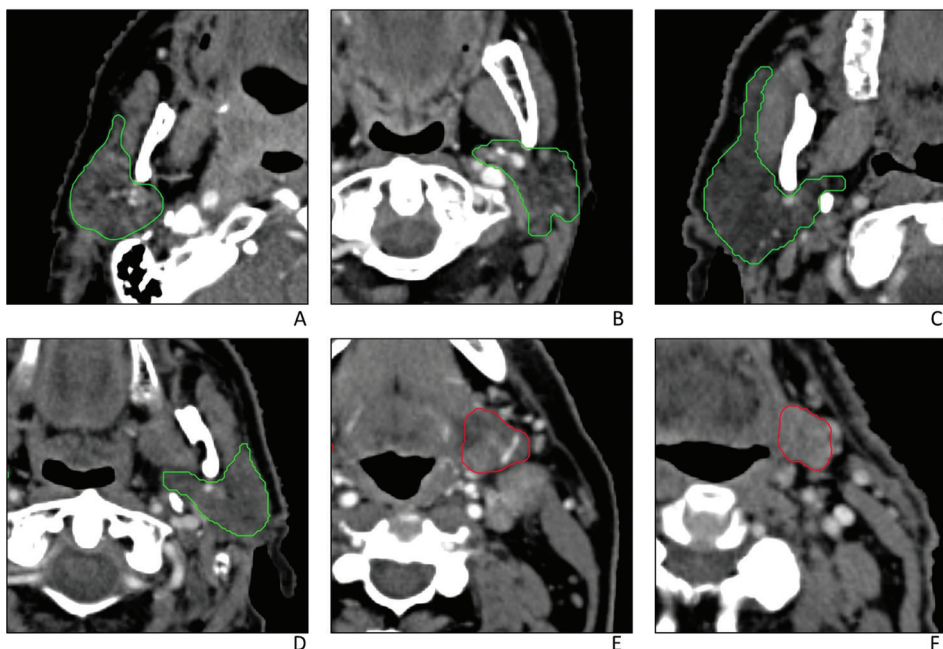


Figure 2 Examples of patients with high (A-B) and low (C-D) Short Run Emphasis values of the parotid gland. Examples of submandibular glands with high (E) and low (F) maximum CT intensity value.

Maximum Intensity and sticky saliva

Our multivariable analysis showed that the maximum CT intensity value of the submandibular gland was associated with $STIC_{12m}$. This maximum CT intensity was related to intra-vascular contrast in the artery or vein supplying the submandibular gland (figure 2E-F). There are no studies reported that support our finding that there is a relationship between vascularisation of the submandibular gland and the development of sticky saliva. Both lasso and internal bootstrapped validation showed robust improvement of prediction with the maximum intensity. However, this IBM was not very stable for the inter-observer variation in delineations of the submandibular glands. Since the blood vessels supplying the submandibular gland can be located at the border of the gland, they are not always delineated, resulting in this marginal stability. Additionally, we expect that the timing of, or the absence of intravenous contrast admitted during acquisition will have a big impact on this IBM. This IBM seems, therefore, suboptimal and



further research is necessary to investigate whether there is an underlying mechanism. For example, higher perfusion could relate to higher oxidation of the submandibular gland, thus increasing the radio-sensitivity. Furthermore, the significant improvement of the prediction of $STIC_{12m}$ by the maximum CT intensity of the submandibular gland should be tested in an external dataset.

Robustness of modeling

The risk of finding false positive associations and overfitting the model were partly addressed by pre-selecting variables based on their inter-correlation. Additionally, we performed alternative multivariable analyses, including logistic regression with forward and backward variable selection based on log-likelihood and the Akaike information criterion (AIC), respectively. The dominating factors selected by these analyses were the same as selected by the lasso regularisation. The same was true if forward selection was performed without pre-selection. Therefore, the selected variables were independent of the method of analysis. This suggests the stability of the associations in this dataset are relatively high. Furthermore, coefficients and performance measures of all models were corrected for optimism by means of internal validation. However, the model selection procedure was not included in the internal validation, as this inhibited model comparison, and so further external validation is warranted.

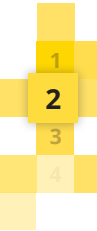
Clinical impact

In this study was shown that the NTCP models based on dose and baseline complaints were significantly improved with IBMs. Nevertheless, the clinical impact of the model improvement in terms of classification and performance remains limited at this point in time. Yet we consider the current study important, as it is an initial step to improve understanding of the patient-specific response of healthy tissue to RT, hereby leading to better identification of HNC patients at risk of developing side effects.

Conclusion

Prediction of xerostomia and sticky saliva 12 months after RT was significantly improved by including CT characteristics of the parotid and submandibular glands for our patient group. The CT image biomarker that positively associated with higher probability of developing xerostomia was “short run emphasis”,

which might be a measure of non-functional fatty parotid tissue. The maximum CT intensity in the submandibular glands was associated with sticky saliva, and probably related with vascularization. These image biomarkers are a first step to identifying patient characteristics that explain the patient-specific response of healthy tissue to dose.



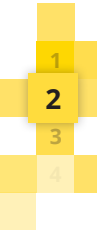
References

- 1 Pignon JP, Maître A le, Maillard E, Bourhis J. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): An update on 93 randomised trials and 17,346 patients. *Radiother Oncol* 2009;92:4-14.
- 2 Bonner JA, Harari PM, Giralt J, Azarnia N, Shin DM, Cohen RB, et al. Radiotherapy plus Cetuximab for Squamous-Cell Carcinoma of the Head and Neck. *N Engl J Med* 2006;354:567-78.
- 3 Machtay M, Moughan J, Trotti A, Garden AS, Weber RS, Cooper JS, et al. Factors associated with severe late toxicity after concurrent chemoradiation for locally advanced head and neck cancer: An RTOG analysis. *J Clin Oncol* 2008;26:3582-9.
- 4 Beetz I, Schilstra C, Van Der Schaaf A, Van Den Heuvel ER, Doornaert P, Van Luijk P, 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. *Radiother Oncol* 2012;105:101-6.
- 5 Jellema AP, Doornaert P, Slotman BJ, Leemans CR, Langendijk J a. Does radiation dose to the salivary glands and oral cavity predict patient-rated xerostomia and sticky saliva in head and neck cancer patients treated with curative radiotherapy? *Radiother Oncol* 2005;77:164-71.
- 6 Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5.
- 7 Marzi S, Pinnarò P, D'Alessio D, Strigari L, Bruzzaniti V, Giordano C, et al. Anatomical and dose changes of gross tumour volume and parotid glands for head and neck cancer patients during intensity-modulated radiotherapy: effect on the probability of xerostomia incidence. *Clin Oncol (R Coll Radiol)* 2012;24:e54-62.
- 8 Bronstein a D, Nyberg D a, Schwartz a N, Shuman WP, Griffin BR. Increased salivary gland density on contrast-enhanced CT after head and neck radiation. *AJR Am J Roentgenol* 1987;149:1259-63.
- 9 Teshima K, Murakami R, Tomitaka E, Nomura T, Toya R, Hiraki A, et al. Radiation-induced parotid gland changes in oral cancer patients: correlation between parotid volume and saliva production. *Jpn J Clin Oncol* 2010;40:42-6.
- 10 Nishimura Y, Nakamatsu K, Shibata T, Kanamori S, Koike R, Okumura M, et al. Importance of the initial volume of parotid glands in xerostomia for patients with head and neck cancers treated with IMRT. *Jpn J Clin Oncol* 2005;35:375-9.
- 11 Scalco E, Fiorino C, Cattaneo GM, Sanguineti G, Rizzo G. Texture analysis for the assessment of structural changes in parotid glands induced by radiotherapy. *Radiother Oncol* 2013;109:384-7.

- 12 Brouwer CL, Steenbakkens RJHM, Bourhis J, Budach W, Grau C, Grégoire V, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother Oncol* 2015;117:83–90.
- 13 van der Laan HP, Christianen MEMC, Bijl HP, Schilstra C, Langendijk J a. The potential benefit of swallowing sparing intensity modulated radiotherapy to reduce swallowing dysfunction: an in silico planning comparative study. *Radiother Oncol* 2012;103:76–81.
- 14 Christianen MEMC, van der Schaaf A, van der Laan HP, Verdonck-de Leeuw IM, Doornaert P, Chouvalova O, et al. Swallowing sparing intensity modulated radiotherapy (SW-IMRT) in head and neck cancer: Clinical validation according to the model-based approach. *Radiother Oncol* 2015.
- 15 Christianen MEMC, Langendijk JA, Westerlaan HE, Van De Water TA, Bijl HP. Delineation of organs at risk involved in swallowing for radiotherapy treatment planning. *Radiother Oncol* 2011;101:394–402.
- 16 Grégoire V, Levendag P, Ang KK, Bernier J, Braaksma M, Budach V, et al. CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC, RTOG consensus guidelines. *Radiother Oncol* 2003;69:227–36.
- 17 Christianen MEMC, Schilstra C, Beetz I, Muijs CT, Chouvalova O, Burlage FR, et al. Predictive modelling for swallowing dysfunction after primary (chemo)radiation: results of a prospective observational study. *Radiother Oncol* 2012;105:107–14.
- 18 Beetz I, Schilstra C, Burlage FR, Koken PW, Doornaert P, Bijl HP, 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. *Radiother Oncol* 2012;105:86–93.
- 19 Vergeer MR, Doornaert PAH, Rietveld DHF, Leemans CR, Slotman BJ, Langendijk JA. Intensity-modulated radiotherapy reduces radiation-induced morbidity and improves health-related quality of life: results of a nonrandomized prospective study using a standardized follow-up program. *Int J Radiat Oncol Biol Phys* 2009;74:1–8.
- 20 Houweling AC, Philippens MEP, Dijkema T, Roesink JM, Terhaard CHJ, Schilstra C, et al. A comparison of dose-response models for the parotid gland in a large group of head-and-neck cancer patients. *Int J Radiat Oncol Biol Phys* 2010;76:1259–65.
- 21 Haralick R, Shanmugan K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973;3:610–21.
- 22 Tang X. Texture information in run-length matrices. *IEEE Trans Image Process* 1998;7:1602–9.
- 23 Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, et al. Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification. *Pattern Recognit Inf Process* 2009:140–5.
- 24 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;57:289–300.



- 25 Van Der Schaaf A, Xu CJ, Van Luijk P, Van'T Veld A a., Langendijk J a., Schilstra C. Multivariate modeling of complications with data driven variable selection: Guarding against overfitting and effects of data set size. *Radiother Oncol* 2012;105:115–21.
- 26 Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010;33.
- 27 Hesterberg T, Choi NH, Meier L, Fraley C. Least angle and L1 penalized regression: A review. *Stat Surv* 2008;2:61–93.
- 28 Roberts S, Nowak G. Stabilizing the lasso against cross-validation variability. *Comput Stat Data Anal* 2014;70:198–211.
- 29 Steyerberg EW, Harrell FE, Borsboom GJJ., Eijkemans MJ., Vergouwe Y, Habbema JDF. Internal validation of predictive models. *J Clin Epidemiol* 2001;54:774–81.
- 30 Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography (Cop)* 2013;36:027–46.
- 31 R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: the R Foundation for Statistical Computing. 2011:Available online at <http://www.R-project.org/>.
- 32 Brouwer CL, Steenbakkers RJ, van den Heuvel E, Duppen JC, Navran A, Bijl HP, et al. 3D Variation in delineation of head and neck organs at risk. *Radiat Oncol* 2012;7:32.
- 33 Langendijk JA, Lambin P, De Ruyscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. *Radiother Oncol* 2013;107:267–73.
- 34 Izumi M, Hida a, Takagi Y, Kawabe Y, Eguchi K, Nakamura T. MR imaging of the salivary glands in sicca syndrome: comparison of lipid profiles and imaging in patients with hyperlipidemia and patients with Sjogren's syndrome. *AJR AmJRoentgenol* 2000;175:829–34.
- 35 Burke CJ, Thomas RH, Howlett D. Imaging the major salivary glands. *Br J Oral Maxillofac Surg* 2011;49:261–9.
- 36 Broggi S, Fiorino C, Dell'Oca I, Dinapoli N, Paiusco M, Muraglia A, et al. A two-variable linear model of parotid shrinkage during IMRT for head and neck cancer. *Radiother Oncol* 2010;94:206–12.



Chapter 3

^{18}F -FDG PET image biomarkers improve prediction of late radiation-induced xerostomia

Published in: **Radiotherapy and Oncology** 2017 September;126:89–95

van Dijk LV, Noordzij W, Brouwer CL, Boellaard R, Burgerhof JGM, Langendijk JA, Sijtsema NM, Steenbakkers RJHM

Online Supplemental Materials : <https://doi.org/10.1016/j.radonc.2017.08.024>

Abstract

Background and purpose

Current prediction of radiation-induced xerostomia 12 months after radiotherapy ($\text{Xer}_{12\text{m}}$) is based on mean parotid gland dose and baseline xerostomia ($\text{Xer}_{\text{baseline}}$) scores. The hypothesis of this study was that prediction of $\text{Xer}_{12\text{m}}$ is improved with patient-specific characteristics extracted from ^{18}F -FDG PET images, quantified in PET image biomarkers (PET-IBMs).

Materials and Methods

Intensity and textural PET-IBMs of the parotid gland were collected from pre-treatment ^{18}F -FDG PET images of 161 head and neck cancer patients. Patient-rated toxicity was prospectively collected. Multivariable logistic regression models resulting from step-wise forward selection and Lasso regularisation were internally validated by bootstrapping. The reference model with parotid gland dose and $\text{Xer}_{\text{baseline}}$ was compared with the resulting PET-IBM models.

Results

High values of the intensity PET-IBM (90th percentile (P90)) and textural PET-IBM (Long Run High Gray-level Emphasis 3 (LRHG3E)) were significantly associated with lower risk of $\text{Xer}_{12\text{m}}$. Both PET-IBMs significantly added in the prediction of $\text{Xer}_{12\text{m}}$ to the reference model. The AUC increased from 0.73 (0.65-0.81) (reference model) to 0.77 (0.70-0.84) (P90) and 0.77 (0.69-0.84) (LRHG3E).

Conclusion

Prediction of $\text{Xer}_{12\text{m}}$ was significantly improved with pre-treatment PET-IBMs, indicating that high metabolic parotid gland activity is associated with lower risk of developing late xerostomia. This study highlights the potential of incorporating patient-specific PET-derived functional characteristics into NTCP model development.

Introduction

^{18}F -FDG PET imaging provides functional information about the metabolic activity of tissue. This makes ^{18}F -FDG PET a powerful and widely used diagnostic modality in oncology. In head and neck oncology, ^{18}F -FDG PET can complement other image modalities in tumour staging and delineation for radiotherapy [1,2]. The common clinical use of ^{18}F -FDG PET allows for the possibility to extract large amounts of patient-specific functional information that could contribute to prognosis for head and neck cancer (HNC) patients. Several studies have shown that PET image characteristics of the tumour can contribute to predicting overall, disease-free or event-free survival [3–6]. However, patient-specific image characteristics for predicting normal tissue radiation toxicities are less explored, while these are also crucial in supporting treatment decisions. Additionally, new radiation techniques (e.g. proton therapy [7] and Magnetic Resonance Imaging (MRI) guided radiation [8]) may allow for better sparing of normal tissue. These new techniques demand improved prediction models, to select patients most at risk of developing toxicities[9].

Radiation-induced xerostomia is a major and frequent side effect for HNC patients, and has a considerable impact on these patients' quality of life [10]. Conventional Normal Tissue Complication Probability (NTCP) models that predict patient-rated xerostomia are based on dose-volume parameters and baseline complaints [11,12]. However, there is still a significant, unexplained variance in predicting xerostomia with these models. Therefore, the demand persists to improve the identification of patients at risk. Previous work showed that patient-specific CT characteristics of the parotid glands could significantly improve the prediction of patient-rated xerostomia, however, model performance improvement was marginal [13]. The hypothesis was that the predictive CT characteristic is related to the ratio of non-function to functional parotid tissue. It can be expected that this ratio would be better represented by image characteristics from functional imaging (i.e. PET or MR images).

In this study, the relationship was tested between metabolic activity of the parotid gland and late xerostomia. Consequently, the patient-specific response to radiation in developing this toxicity was investigated. The purpose was to determine whether functional information from ^{18}F -FDG PET images, which is quantified in PET-image biomarkers (PET-IBMs), was associated with patient-rated moderate-to-severe xerostomia 12 months after radiotherapy ($\text{Xer}_{12\text{m}}$). Since



current NTCP prediction models are based on parotid gland dose and baseline complaints, the study subsequently addressed whether PET-IBMs could improve on the current prediction of $\text{Xer}_{12\text{m}}$

Materials and methods

Patient demographics and treatment

^{18}F -FDG PET/CT scans were acquired of 161 HNC patients in treatment position before the start of radiotherapy. The patients were treated with definitive radiotherapy either with or without concurrent chemotherapy or cetuximab, between November 2010 and August 2015. Patients without follow-up data 12 months after radiotherapy were excluded from this study. Patients were also excluded if they underwent surgery in the head and neck area before or within one year after treatment.

A detailed description of the radiotherapy protocols is given in previous studies [13,14]. In summary, all patients were treated with IMRT or VMAT using a simultaneous integrated boost (SIB) technique. The parotid glands and the swallowing structures were spared as much as possible without compromising the dose to the target volumes [14,15]. Patients received a total dose of 70 Gy (2 Gy per fraction, 5 or 6 times a week) to the primary tumour and, if present, pathological lymph nodes. A radiation dose of 54.25 Gy (1.55 Gy per fraction, 5 or 6 times a week) was delivered to the elective lymph node levels.

Endpoints

The primary endpoint was patient-rated moderate-to-severe xerostomia 12 months after radiotherapy ($\text{Xer}_{12\text{m}}$), which corresponds to the 2 highest scores of the 4-point Likert scale of the EORTC QLQ-H&N35 questionnaire. This endpoint was prospectively assessed as part of a Standard Follow-up Program (SFP) for Head and Neck Cancer Patients (NCT02435576), as described in previous studies [11,12,16].

Dose and clinical parameters

For treatment planning, parotid glands were delineated on the planning (PET/CT) scans. The mean dose to both the contra- and ipsilateral parotid and submandibular glands were extracted from the dose-volume information [11,17].

In addition, baseline patient-rated xerostomia ($Xer_{baseline}$) was also considered (none vs. any).

Patient characteristics such as age, sex, WHO-performance, tumour stage and body mass index did not significantly add to the parotid gland dose and $Xer_{baseline}$ in predicting Xer_{12m} in previous studies [11,13,18]. This was again observed in the current cohort, therefore these variables were not further reported in this study.

¹⁸F-FDG PET acquisition

Approximately 2 weeks before the start of radiotherapy, ¹⁸F-FDG PET/CT images (Siemens Biograph 64-slice PET/CT scanner, Siemens Medical Systems, Knoxville, TN, USA) were acquired in with the patient positioned for radiotherapy. PET/CT system performance were initially harmonized conform the Netherlands protocol for FDG PET imaging [19] and later by EARL accreditation [20].

Patients were instructed not to eat or drink 6 hours before scanning, but were encouraged to drink water to ensure adequate hydration. A body weight-based intravenous injection dose of 3 MBq/kg was administered 60 minutes prior to the ¹⁸F-FDG PET acquisition. ¹⁸F-FDG PET images were acquired in the caudal–cranial direction with an acquisition time of ~3 min per bed position.

Candidate PET-image biomarkers

Intensity PET-IBMs were extracted, representing first order standardized uptake value (SUV) characteristics of the delineated contra-lateral parotid glands. Examples are mean, minimum, maximum, standard deviation and root mean square of the SUVs. For the complete list of the 24 intensity PET-IBMs, see supplementary data 1. Figure 1 shows a schematic representation of PET- IBMs extraction process.

Furthermore, more complex, textural features were extracted describing the intensity heterogeneity. These textural PET-IBMs were extracted from the grey level co-occurrence matrix (GLCM) [21], grey level run-length matrix (GLRLM) [22,23], grey level size-zone matrix (GLSZM) [24] and neighbourhood grey tone difference matrix (NGTDM) [25]. GLCM describes the grey level transitions, GLRLM and GLSZM describe the directional and volumetric grey level repetitions, respectively. NGTDM describes the relationship of sum and averages of grey level differences of direct adjacent voxels.

For this study, the average of PET-IBMs from GLCM and GLRLM in 13 independent directions were used. The range of SUVs were binned with a fixed bin size of 0.25.



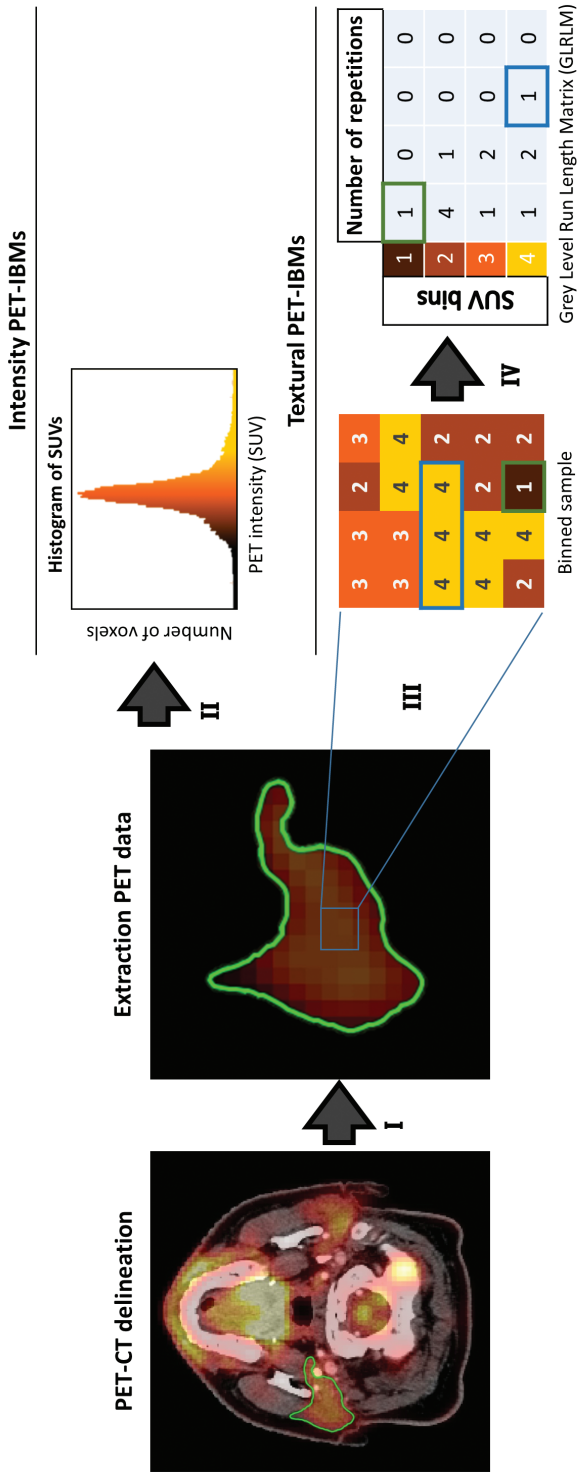


Figure 1 Example of PET-IBM extraction process. The PET information from the delineated parotid gland was extracted (I). Intensity PET-IBMs were obtained from all voxels inside this contour (II). The SUVs were binned for the textural analysis (III). For illustration, a Grey Level Run Length Matrix is constructed from a binned sample, it quantifies the number of repetitions of binned SUVs from left to right (IV).

Discretization of SUV is necessary to reduce the number of possible intensity values, and so reduce noise when calculating textural features [26]. All 66 textural PET-IBMs (25 GLCM, 18 GLRLM, 18 GLSZM and 5 NGTDM) were normalized by subtracting the average from the PET-IBMs values and then dividing by the standard deviation. For the complete list refer to supplementary data 2. All PET-IBMs were extracted in MATLAB (version R2014a).

Univariable analysis

Univariable logistic regression analysis was performed to evaluate the basic associations of PET-IBMs with late xerostomia. P-values < 0.05 were considered statistically significant. Coefficients (β) were evaluated to understand the effect that is described by the PET-IBMs in relation to Xer_{12m} . The univariable analysis was not used for the variable selection.

Multivariable analysis

Reference model

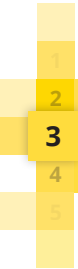
A reference prediction model was evaluated for the current patient cohort. This model was based on the mean dose to the contralateral parotid gland and $Xer_{baseline}$. These were the predictors that were identified by Beetz et al.[11].

Intensity and textural PET-IBMs

First, a basic PET-IBM model was created by adding the 'mean SUV' of the parotid gland as an extra variable to the reference model. Since this variable is the simplest of PET-IBMs, it is the easiest to interpret.

Both step-wise forward selection and Lasso regularisation were performed for multivariable logistic analysis of the PET-IBMs, together with parotid dose and $Xer_{baseline}$. Step-wise forward selection was based on the largest significant log-likelihood differences [27]. Lasso regularisation uses the penalisation term lambda, which excludes variables by reducing their coefficients to zero. The optimal lambda was determined by 100-times repeated 10-fold cross validation [28].

To understand the contribution of the different types of PET-IBMs to the reference model, the model analysis of all SUV intensity and textural PET-IBMs were conducted separately. Subsequently, the resulting SUV intensity and textural models were compared to the reference and the 'mean SUV' model. The performance of the constructed models was quantified with the Area Under the



ROC curve (AUC), the Nagelkerke R^2 and the discrimination slope. Furthermore, calibration was evaluated with the Hosmer–Lemeshow test. Internal validation was performed with bootstrapping to correct for optimism of the model [29,30]. Analyses were performed with the R-packages ‘Lasso and Elastic-Net Regularized Generalized Linear Models’ (version 2.0-2)[28] and ‘Regression Modeling Strategies’ (version 4.3-1)[31].

Inter-variable relationships

The relationship between variables of predictive PET-IBMs (and $\text{Xer}_{\text{baseline}}$) was investigated with Pearson correlation (continuous variables) and univariable logistic regression analysis (binary variables). Furthermore, in a previous study, the short run emphasis (SRE), which was extracted from CT information of the parotid gland, was significantly associated with $\text{Xer}_{12\text{m}}$ [13]. In the current study, this SRE was also extracted from the CT-scans of patients without metal artefacts in the images. Subsequently the correlation of the CT-based SRE values and the predictive PET-IBMs was tested. Additionally, the improvement of the PET-IBM or reference models by SRE was also tested in this patient subset.

Results

Patients

Patient characteristics are depicted in Table 1. Briefly, nearly all patients were bi-laterally irradiated, most patients had oropharyngeal carcinomas and had no baseline xerostomia (none vs. any: 61% vs. 39%). Sixty of the 161 (37%) patients developed moderate-to-severe xerostomia ($\text{Xer}_{12\text{m}}$).

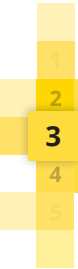
Univariable analysis

In the univariable analysis, the mean dose to the parotid gland and $\text{Xer}_{\text{baseline}}$ were both associated with $\text{Xer}_{12\text{m}}$. Univariable analysis showed that 11 of 24 intensity PET-IBMs and 35 of 66 textural PET-IBMs were significantly associated with $\text{Xer}_{12\text{m}}$ (supplementary data 3). In general, a negative coefficient was observed for PET-IBMs that have a positive relationship with SUVs in the parotid gland, indicating that low parotid gland SUVs were associated with a high $\text{Xer}_{12\text{m}}$ risk.

Table 1 Patient characteristics

Characteristics	N=161	%
Sex		
Female	50	31
Male	111	69
Age		
18 - 65 years	95	59
> 65 years	66	41
Tumour site		
Oropharynx	78	48
Nasopharynx	7	4
Hypopharynx	18	11
Larynx	51	32
Oral cavity	7	4
Tumour classification		
T1	14	9
T2	51	32
T3	52	32
T4	44	27
Node classification		
N0	71	44
N1	14	9
N2abc	74	46
N3	2	1
Systemic treatment		
yes	71	44
no	90	56
Treatment technique		
IMRT	145	90
VMAT	16	10
Bi-lateral		
yes	139	86
no	22	14
Baseline Xerostomia		
No	98	61
A bit	46	29
Quite a bit	13	8
A lot	4	2

Abbreviations: IMRT: Intensity-Modulated Radiation Therapy; VMAT: Volumetric Arc Therapy



Multivariable analysis

Reference model

The reference model with the variables contra-lateral parotid gland dose and $Xer_{baseline}$ (none vs. any) was fit to the dataset (Table 2). The performance measures are depicted in Table 3 (AUC =0.73 (0.65-0.81), $R^2=0.22$).

Intensity PET-IBMs

First, the basic PET-IBM model ('mean SUV', parotid dose, $Xer_{baseline}$) showed that the addition of the 'mean SUV' significantly improved the reference model (Likelihood ratio test; $p=0.005$). Consistent with the univariable analysis, the negative regression coefficient of the mean SUV indicates that high mean SUVs were associated with a lower Xer_{12m} risk (Table 2). The performance of this basic PET-IBM model (AUC =0.77 (0.69-0.84), $R^2=0.27$), was better than that of the reference model (Table 3).

Resulting from both the Lasso regularisation and forward selection, the 90th percentile of SUVs (P90) was the most predictive of all intensity PET-IBMs (Figure 2), leading to a significant (Likelihood-ratio test; $p=0.002$), substantial improvement of the model performance measures (Table 2 and 3; AUC=0.77 (0.70-0.84), $R^2=0.28$) compared to the reference model (AUC =0.73 (0.64-0.83), $R^2=0.23$). High correlations were observed between P90 and the IBMs that could also significantly improve the reference model when individually added to the reference model ($\rho = 0.82 \pm 0.15$). See supplementary 4 for the correlations of PET-IBMs.

In Figure 3 the NTCP curves for different P90 values are depicted of the following P90 model:

$$NTCP = \frac{1}{1 - e^{-s}}$$

where

$$s = 0.984 + 0.048 \cdot \text{Contra Dose (PG)} + 1.402 \cdot Xer_{baseline} - 1.527 \cdot P90(PG)$$

Textural PET-IBMs

The most predictive textural PET-IBM was the Long Run High Gray-level Emphasis 3 (LRHG3E), which is derived from the GLRLM. The value of this PET-IBM increases when long repetitions of high SUVs are present in the parotid gland with extra (power of 3) emphasis on high SUVs (see supplementary data 2 for formula). This variable was selected by both the Lasso regularisation and the step-wise

forward selection. This variable significantly improved the reference model in predicting Xer_{12m} (Likelihood-ratio test; $p=0.001$). The negative coefficient of LRHG3E indicated once more that high SUVs are associated with low Xer_{12m} risk (Table 2). The addition of LRHG3E improved the reference model performance (0.77 (0.69-0.84), $R^2=0.29$; Table 3). The NTCP curves for different LRHG3E are depicted in Figure 3 for the following model:

$$NTCP = \frac{1}{1 - e^{-s}}$$

where

$$s = -2.598 + 0.051 \cdot \text{Contra Dose (PG)} + 1.479 \cdot Xer_{baseline} - 0.880 \cdot \frac{LRHG3E (PG) - 201.24}{177.05}$$

Table 2 Estimated coefficients (uncorrected and corrected for optimism) of reference model and PET-IBMs model

	β		OR (95% CI)	p-value
	Uncorrected	Corrected		
<i>intercept</i>	-2.633	-2.579		
<i>Xer_{baseline}</i>	1.559	1.526	4.75 (2.32-9.75)	<0.001
<i>PG dose</i>	0.056	0.054	1.06 (1.02-1.10)	0.002
<i>intercept</i>	0.906	0.828		
<i>Xer_{baseline}</i>	1.473	1.384	4.36 (2.08-9.14)	<0.001
<i>PG dose</i>	0.051	0.047	1.05 (1.01-1.09)	0.007
Mean SUV	-1.776	-1.669	0.17 (0.05-0.64)	0.009
<i>Intercept</i>	1.070	0.984		
<i>Xer_{baseline}</i>	1.487	1.402	4.43 (2.10-9.31)	<0.001
<i>PG dose</i>	0.050	0.048	1.05 (1.01-1.09)	0.007
P90	-1.620	-1.527	0.20 (0.06-0.63)	0.006
<i>intercept</i>	-2.752	-2.598		
<i>Xer_{baseline}</i>	1.577	1.479	4.84 (2.29-10.22)	<0.001
<i>PG dose</i>	0.055	0.051	1.05 (1.02-1.10)	0.004
LRHG3E	-0.938	-0.880	0.39 (0.19-0.82)	0.013

Abbreviations: $Xer_{baseline}$: xerostomia at baseline; PG dose: contralateral mean dose to parotid gland; P90: 90th percentile of intensities; LRHG3E: Long Run High Gray-level Emphasis 3; β : regression coefficients; OR: odds ratio; CI: confidence interval

Inter-variable relationships

The predictive PET-IBMs P90 (intensity) and LRHG3E (textural) were closely correlated ($p<0.001$; $r=0.83$). Moreover, they did not add independent information

to each other in predicting $\text{Xer}_{12\text{m}}$ (Likelihood ratio test; $p > 0.21$). Univariable logistic analysis showed no significant association between $\text{Xer}_{\text{baseline}}$ and P90 ($p = 0.079$) or LRHG3E ($p = 0.465$).

In the current study cohort, 100 patients did not have metal artefacts in the CT images and could therefore be used for the analysis of the CT-based IBM, the short run emphasis (SRE)[13]. This CT-based SRE was significantly correlated to the predictive PET-IBMs P90 ($p = 0.008$; $r = -0.26$) and LRHG3E ($p = 0.026$; $r = -0.22$). The SRE neither significantly improved the reference model (likelihood ratio test, $p = 0.055$), nor did it add to the PET-IBM models with P90 (likelihood ratio test, $p = 0.140$) and LRHG3E (likelihood ratio test, $p = 0.096$) in this cohort subset.

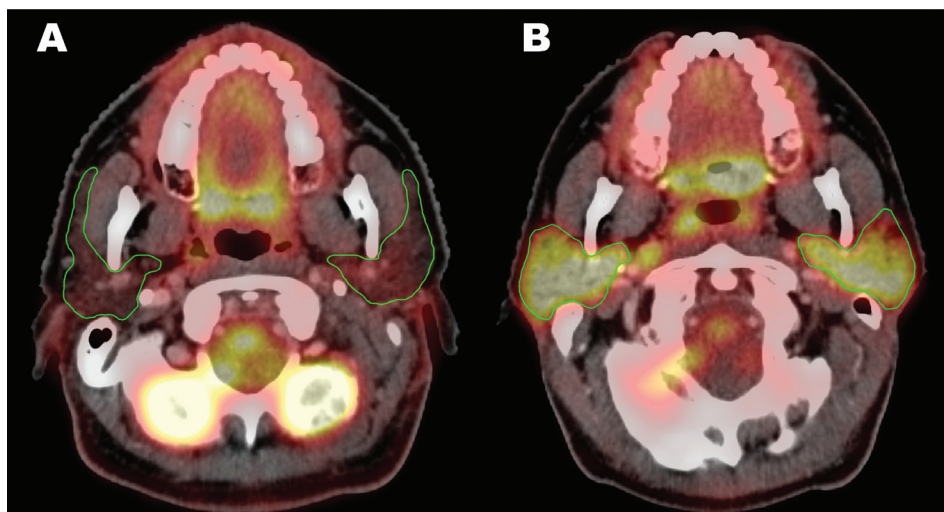


Figure 2 Example of patients with A) low and B) high values of mean SUV, P90 and LRHG3E, which were associated with A) higher and B) lower risk of developing $\text{Xer}_{12\text{m}}$. Scaling in both images: 0.5 to 3.5 SUV

Discussion

This study is novel to show that the high metabolic activity of the parotid gland was associated with a lower risk of developing late xerostomia ($\text{Xer}_{12\text{m}}$). Moreover, the prediction of late xerostomia was significantly and substantially improved with addition of patient-specific PET-IBMs to the reference model based on dose and $\text{Xer}_{\text{baseline}}$. These findings could improve understanding of normal tissue response following radiotherapy, since the variation in patient-specific PET characteristics can partly explain the unexplained variance in predicting

Table 3 Performance of NTCP models with and without PET-IBMs

	Reference model		PET-IBMs models			
	Xerbaseline PG dose	Xerbaseline PG dose	Xerbaseline PG dose	Xerbaseline PG dose	Xerbaseline PG dose	Xerbaseline PG dose
Overall	184.51	176.57	175.30	174.31	174.31	174.31
	0.22	0.27	0.28	0.29	0.29	0.29
Area Under the Curve (AUC)	0.73 (0.65-0.81)	0.77 (0.69-0.84)	0.77 (0.70-0.84)	0.77 (0.69-0.84)	0.77 (0.69-0.84)	0.77 (0.69-0.84)
Discrimination slope	0.17	0.20	0.21	0.21	0.21	0.21
HL test X ² (p-value)	11.22 (0.19)	4.24 (0.83)	6.72 (0.57)	6.30 (0.61)	6.30 (0.61)	6.30 (0.61)
Calibration slope (intercept)	1.00 (0.00)	0.95 (0.02)	0.95 (0.02)	0.99 (0.00)	0.99 (0.00)	0.99 (0.00)
AUC _{corrected}	0.72	0.75	0.76	0.75	0.75	0.75
Nagelkerke R ² _{corrected}	0.20	0.24	0.25	0.26	0.26	0.26

HL: Hosmer-Lemeshow; corrected: corrected for optimism with bootstrapping; IBM: Image Biomarker; Xer_{baseline}: xerostomia at baseline; PG dose: contralateral mean dose to parotid gland; P90: 90th percentile of intensities; LRHG3E: Long Run High Gray-level Emphasis 3



xerostomia with dose parameters. Moreover, it could improve identification of patients that are at risk of late radiation-induced side effects, which could potentially benefit most from new therapy technology such as proton [7] and MRI-guided irradiation [8]. In other words, better prediction of toxicities could improve the treatment decision support [9,32]. However, external validation of the PET-IBM models in an independent dataset is necessary before clinical implementation [33].

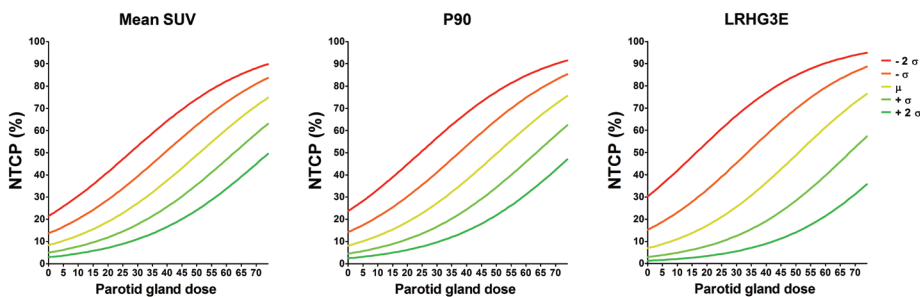


Figure 3 Normal Tissue Complication Probability (NTCP) values for late xerostomia (Xer_{12m}) of models based on mean SUV(left), P90 (middle) and LRHG3E (right). Curves are given for the mean PET-IBM values (P90: $\mu=2.23$; LRHG3E: $\mu=201.24$) and for 1 and 2 standard deviation from these mean values (mean SUV: $\mu=1.93$, $\sigma=0.33$; P90: $\mu= 2.23$, $\sigma=0.41$; LRHG3E: $\mu= 201.24$, $\sigma=177.05$). For these curves no baseline xerostomia was assumed.

The PET-IBM that indicates the minimum value of the 90% highest SUVs (P90) was the most predictive of all intensity PET-IBMs. The mean SUV also performed well, but P90 appeared more relevant in this dataset. A high P90 was associated with a lower risk of developing late xerostomia. Similar effect and predictive improvement was observed from LRHG3E (Long Run High Gray-level Emphasis 3) of the textural PET-IBMs, which significantly correlated with P90 ($\rho=0.83$). This PET-IBM indicates high SUVs that are spatially adjacent to each other. Both PET-IBMs were negatively associated with Xer_{12m} , suggesting that patients with low metabolic activity in the parotid glands were at risk of developing late xerostomia. Although both P90 and LRHG3E perform similarly, currently the P90 is simpler to calculate. However, LRHG3E also contains information about the spatial connectivity of the high SUV voxels, i.e. large repetitions of voxels with high SUV increase the LRHG3E values. External validation is needed to confirm the predictive power of LRHG3E over P90. Additionally, an alternative variable selection approach, Lasso regularization, resulted in very comparable final

models. Since they were independent of the method of analysis, it suggests that the associations in this dataset were relatively stable.

Predictive PET-IBMs were not significantly associated with $Xer_{baseline}$. This suggests that PET-IBMs contain unique and additional information to baseline xerostomia complaints, since the addition of PET-IBMs to $Xer_{baseline}$ (and PG dose) improved the prediction of Xer_{12m} significantly.

This study suggests that high metabolic parotid glands have more viable cells (parenchyma and/or stem cells) with more repair capability and/or are less radiosensitive. Although possibly driven by multiple underlying biological processes, there is some similarity in the tumour reaction to radiation. For tumour tissue it is known that high metabolic tumours are more likely to recur [34], particularly in their high metabolic regions [35]. A possible explanation is that it arises from a combination of higher cell density, proliferation rate of metabolically active tissue and DNA repair capacity [36].

Other studies have shown that parotid gland SUVs decrease post-radiotherapy, and in addition that this change was associated with parotid gland dose [37,38]. Cannon et al. [38] showed that mean 'SUV-weighted parotid gland dose (voxel-wise)' was significantly related to fractional-SUV (post-SUV/pre-SUV). In an additional small cohort (n=8), they showed that fractional-SUV was significantly associated with fractional salivary flow and physician-rated xerostomia. Although this indirectly suggests that 'SUV-weighted parotid gland dose' is related to xerostomia measures, the direct and separate associations of parotid gland dose and pre-treatment SUV with xerostomia measures or fractional SUV were unfortunately not described.

In previous work, a positive association was shown between higher risk of developing late xerostomia and CT-based SRE (Short Run Emphasis), which might be related to the ratio between non-functional fatty tissue and functional parotid parenchyma tissue. In this study, we showed that this CT-IBM was significantly correlated to P90 and LRGH3E in patients without metal artefacts (n=100) and did not significantly add to the PET-IBM models. Additionally, the performance of predicting Xer_{12m} was substantially higher with PET-based IBM models than with CT-based IBMs. This suggests that ^{18}F -FDG PET is better to quantify the ratio between fatty non-functional and functional parotid parenchyma tissue. This is logical since ^{18}F -FDG PET is a functional image modality. Furthermore, the SRE did not show a significant improvement in the reference model for the



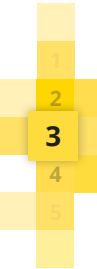
cohort subset, which might be caused by the small additive effect of SRE and low number of patients on which this IBM could be tested.

A well-defined protocol was used to ensure optimal standardisation of SUV in the ^{18}F -FDG PET images by correcting for bodyweight, injection dose, tracer uptake period, and glucose plasma levels by letting the patients fast [19,20]. Although SUVs may also be affected by fasting blood glucose level, muscle activity, liver and kidney function, the images were not corrected for these fluctuations. Furthermore, patients with metal artefacts in CT images were included, where the attenuation correction can influence SUVs, but this bias will primarily be located around the metal implant [20]. Additional analyses showed that the PET-IBMs performance was still good in the sub cohort of patients without metal artefacts. Additionally, future improvements of the consistency and spatial resolution of PET imaging should also improve the performance of the PET-IBMs in predicting $\text{Xer}_{12\text{m}}$.

In this study, patient-rated outcomes (EORTC QLQ-H&N35 questionnaire) were used as a measure for moderate-to-severe xerostomia, because of their relationship with the quality of life of HNC patients [10]. However, some unexplained variability of the models may be caused by the assessment of xerostomia, as the questionnaires can be interpreted differently by the individual patients [39]. Our current study could be strengthened by the addition of investigating the associations between PET-IBMs and objective xerostomia measures. Parotid flow rates are often used, but several studies have shown no or modest correlation between patient reported xerostomia and parotid flow rates [40] and have a low reproducibility [41]. Another example is scintigraphy of parotid gland ejection fraction over time. Although this technique seems promising as a quantitative measure for xerostomia, it requires additional scans with complex procedures with radioactive tracers [41]. This highlights the importance for future research on a non-invasive, accessible and reliable quantitative measure of xerostomia. Nevertheless, we believe that patient-rated xerostomia remains an important endpoint, due to its clinical importance and practical benefits.

Conclusion

The pre-treatment PET-IBMs indicated that a large quantity of high SUVs in the parotid gland was significantly associated with a lower risk of developing xerostomia 12 months after radiotherapy. The addition of the predictive intensity PET-IBM (90th percentile of SUV) to a model with parotid gland dose and baseline xerostomia improved the prediction performance of the reference model substantially (0.73 (0.65-0.81) to 0.77 (0.70-0.84)). This study highlights the importance of incorporating patient-specific functional characteristics into NTCP model development and can, thereby, contribute to the understanding of the patient-specific response of healthy tissue to radiation dose.



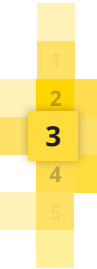
References

- 1 Leclerc M, Lartigau E, Lacornerie T, Daisne JF, Kramar A, Gregoire V. Primary tumor delineation based on (18)FDG PET for locally advanced head and neck cancer treated by chemo-radiotherapy. *Radiother Oncol* 2015;116:87-93.
- 2 Stuckensen T, Kovács AF, Adams S, Baum RP. Staging of the neck in patients with oral cavity squamous cell carcinomas: a prospective comparison of PET, ultrasound, CT and MRI. *J Cranio-Maxillofacial Surg* 2000;28:319-24.
- 3 Abgral R, Keromnes N, Robin P, Le Roux P-Y, Bourhis D, Palard X, et al. Prognostic value of volumetric parameters measured by (18)F-FDG PET/CT in patients with head and neck squamous cell carcinoma. *Eur J Nucl Med Mol Imaging* 2014;41:659-67.
- 4 Eary JF, O'Sullivan F, O'Sullivan J, Conrad EU. Spatial Heterogeneity in Sarcoma 18F-FDG Uptake as a Predictor of Patient Outcome. *J Nucl Med* 2008;49:1973-9.
- 5 Koyasu S, Nakamoto Y, Kikuchi M, Suzuki K, Hayashida K, Itoh K, et al. Prognostic value of pretreatment 18F-FDG PET/CT parameters including visual evaluation in patients with head and neck squamous cell carcinoma. *AJR Am J Roentgenol* 2014;202:851-8.
- 6 Alluri KC, Tahari AK, Wahl RL, Koch W, Chung CH, Subramaniam RM. Prognostic value of FDG PET metabolic tumor volume in human papillomavirus-positive stage III and IV oropharyngeal squamous cell carcinoma. *AJR Am J Roentgenol* 2014;203:897-903.
- 7 Lomax A. Intensity modulation methods for proton radiotherapy. *Phys Med Biol* 1999;44:185-205.
- 8 Langendijk JJW, Raaymakers BW, Raaijmakers AJE, Overweg J, Brown KJ, Kerkhof EM, et al. MRI/linac integration. *Radiother Oncol* 2008;86:25-9.
- 9 Langendijk JA, Lambin P, De Ruyscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. *Radiother Oncol* 2013;107:267-73.
- 10 Langendijk JA, Doornaert P, Verdonck-de Leeuw IM, Leemans CR, Aaronson NK, Slotman BJ. Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. *J Clin Oncol* 2008;26:3770-6.
- 11 Beetz I, Schilstra C, Van Der Schaaf A, Van Den Heuvel ER, Doornaert P, Van Luijk P, 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. *Radiother Oncol* 2012;105:101-6.
- 12 Vergeer MR, Doornaert PAH, Rietveld DHF, Leemans CR, Slotman BJ, Langendijk JA. Intensity-modulated radiotherapy reduces radiation-induced morbidity and improves health-related quality of life: results of a nonrandomized prospective study using a standardized follow-up program. *Int J Radiat Oncol Biol Phys* 2009;74:1-8.
- 13 van Dijk LV, Brouwer CL, van der Schaaf A, Burgerhof JGM, Beukinga RJ, Langendijk JA, et al. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. *Radiother Oncol* 2016:Article in press.

- 14 Van Der Laan HP, Gawryszuk A, Christianen MEMC, Steenbakkers RJHM, Korevaar EW, Chouvalova O, et al. Swallowing-sparing intensity-modulated radiotherapy for head and neck cancer patients: Treatment planning optimization and clinical introduction. *Radiother Oncol* 2013;107:282–7.
- 15 Christianen MEMC, Langendijk JA, Westerlaan HE, Van De Water TA, Bijl HP. Delineation of organs at risk involved in swallowing for radiotherapy treatment planning. *Radiother Oncol* 2011;101:394–402.
- 16 Beetz I, Schilstra C, Burlage FR, Koken PW, Doornaert P, Bijl HP, 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. *Radiother Oncol* 2012;105:86–93.
- 17 Houweling AC, Philippens MEP, Dijkema T, Roesink JM, Terhaard CHJ, Schilstra C, et al. A comparison of dose-response models for the parotid gland in a large group of head-and-neck cancer patients. *Int J Radiat Oncol Biol Phys* 2010;76:1259–65.
- 18 Jellema AP, Doornaert P, Slotman BJ, Leemans CR, Langendijk J a. Does radiation dose to the salivary glands and oral cavity predict patient-rated xerostomia and sticky saliva in head and neck cancer patients treated with curative radiotherapy? *Radiother Oncol* 2005;77:164–71.
- 19 Boellaard R, Oyen WJG, Hoekstra CJ, Hoekstra OS, Visser EP, Willemsen AT, et al. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. *Eur J Nucl Med Mol Imaging* 2008;35:2320–33.
- 20 Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging* 2014;42:328–54.
- 21 Haralick R, Shanmugan K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973;3:610–21.
- 22 Tang X. Texture information in run-length matrices. *IEEE Trans Image Process* 1998;7:1602–9.
- 23 Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process* 1975;4:172–9.
- 24 Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, et al. Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification. *Pattern Recognit Inf Process* 2009:140–5.
- 25 Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern* 1989;19:1264–73.
- 26 Leijenaar RTH, Nalbantov G, Carvalho S, van Elmpt WJC, Troost EGC, Boellaard R, et al. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. *Sci Rep* 2015;5:11075.



- 27 Dehing-Oberije C, De Ruyscher D, Petit S, Van Meerbeeck J, Vandecasteele K, De Neve W, et al. Development, external validation and clinical usefulness of a practical prediction model for radiation-induced dysphagia in lung cancer patients. *Radiother Oncol* 2010;97:455–61.
- 28 Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010;33.
- 29 Van Der Schaaf A, Xu CJ, Van Luijk P, Van 't Veld AA, Langendijk JA, Schilstra C. Multivariate modeling of complications with data driven variable selection: Guarding against overfitting and effects of data set size. *Radiother Oncol* 2012;105:115–21.
- 30 Harrell FE. Regression modeling strategies. With applications to linear models, logistic regression, and survival analysis. 2001.
- 31 R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: the R Foundation for Statistical Computing. 2011:Available online at <http://www.R-project.org/>.
- 32 Kumar V, Gu Y, Basu S, Berglund A, Eschrich S a, Schabath MB, et al. Radiomics: the process and the challenges. *Magn Reson Imaging* 2012;30:1234–48.
- 33 Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *Eur Urol* 2015;67:1142–51.
- 34 Xie P, Yue J-B, Fu Z, Feng R, Yu J-M. Prognostic value of ¹⁸F-FDG PET/CT before and after radiotherapy for locally advanced nasopharyngeal carcinoma. *Ann Oncol* 2010;21:1078–82.
- 35 Due AK, Vogelius IR, Aznar MC, Bentzen SM, Berthelsen AK, Korreman SS, et al. Recurrences after intensity modulated radiotherapy for head and neck squamous cell carcinoma more likely to originate from regions with high baseline [¹⁸F]-FDG uptake. *Radiother Oncol* 2014;111:360–5.
- 36 Jeong J, Deasy JO. Modeling the relationship between fluorodeoxyglucose uptake and tumor radioresistance as a function of the tumor microenvironment. *Comput Math Methods Med* 2014;2014:847162.
- 37 Roach MC, Turkington TG, Higgins K a, Hawk TC, Hoang JK, Brizel DM. FDG-PET assessment of the effect of head and neck radiotherapy on parotid gland glucose metabolism. *Int J Radiat Oncol Biol Phys* 2012;82:321–6.
- 38 Cannon B, Schwartz DL, Dong L. Metabolic imaging biomarkers of postradiotherapy xerostomia. *Int J Radiat Oncol Biol Phys* 2012;83:1609–16.
- 39 Meirovitz A, Murdoch-Kinch CA, Schipper M, Pan C, Eisbruch A. Grading xerostomia by physicians or by patients after intensity-modulated radiotherapy of head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2006;66:445–53.
- 40 Eisbruch A, Kim HM, Terrell JE, Marsh LH, Dawson LA, Ship JA. Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2001;50:695–704.
- 41 Cheng SCH, Wu VWC, Kwong DLW, Ying MTC. Assessment of post-radiotherapy salivary glands. *Br J Radiol* 2011;84:393–402.



Chapter 4

Parotid gland fat related Magnetic Resonance image biomarkers improve prediction of late radiation-induced xerostomia

Published in: **Radiotherapy and Oncology** 2018 September; 128:459–66

van Dijk LV, Thor M, Steenbakkens RJHM, Apte A, Zhai TT, Borra R, Noordzij W, Estilo C, Lee N, Langendijk JA, Deasy JO, Sijtsema NM

Online Supplemental Materials: <https://doi.org/10.1016/j.radonc.2018.06.012>

Abstract

Background and purpose

This study investigated whether Magnetic Resonance image biomarkers (MR-IBMs) were associated with xerostomia 12 months after radiotherapy (Xer_{12m}) and to test the hypothesis that the ratio of fat-to-functional parotid tissue is related to Xer_{12m} . Additionally, improvement of the reference Xer_{12m} model based on parotid gland dose and baseline xerostomia, with MR-IBMs was explored.

Materials and Methods

Parotid gland MR-IBMs of 68 head and neck cancer patients were extracted from pre-treatment T1-weighted MR images, which were normalised to fat tissue, quantifying 21 intensity and 43 texture image characteristics. The performance of the resulting multivariable logistic regression models after bootstrapped forward selection was compared with that of the logistic regression reference model. Validity was tested in a small external cohort of 25 head and neck cancer patients.

Results

High intensity MR-IBM P90 (the 90th intensity percentile) values were significantly associated with a higher risk of Xer_{12m} . High P90 values were related to high fat concentration in the parotid glands. The MR-IBM P90 significantly improved model performance in predicting Xer_{12m} (likelihood-ratio-test; $p=0.002$), with an increase in internally validated AUC from 0.78 (reference model) to 0.83 (P90). The MR-IBM P90 model also outperformed the reference model (AUC=0.65) on the external validation cohort (AUC=0.83).

Conclusion

Pre-treatment MR-IBMs were associated to radiation-induced xerostomia, which supported the hypothesis that the amount of predisposed fat within the parotid glands is associated with Xer_{12m} . In addition, xerostomia prediction was improved with MR-IBMs compared to the reference model.

Introduction

Xerostomia is one of the most frequently reported side-effects following radiotherapy for head and neck cancer, and has a major impact on quality of life [1,2]. Normal Tissue Complication Probability (NTCP) models have been developed to predict radiation-induced xerostomia and have demonstrated a clear relationship with parotid gland dose and baseline patient-rated xerostomia [3,4]. Nevertheless, substantial unexplained variance in predicting xerostomia remains. Better understanding of the aetiology of radiation-induced xerostomia is necessary to advance towards more individualized treatments and better sparing of normal tissues by further dose optimization, by means of new radiation techniques, such as proton therapy [5,6] and Magnetic Resonance Imaging (MRI) guided radiation [7].

Tumour-based image biomarkers (IBMs), which are shape, intensity and texture characteristics extracted from images, can contribute to the prediction of overall, disease-free and progression-free survival [8–13]. However, the role of these IBMs in normal tissues to predict radiation-induced toxicities is less explored, while these are imperative in supporting treatment decisions [5].

Our previous study based on IBMs from pre-treatment CT images, demonstrated that high heterogeneous parotid gland tissue, was associated with a higher probability of developing late xerostomia [14]. Qualitative evaluation of the parotid glands suggested that the predictive CT-IBM indicated the ratio between fatty and functional parotid parenchyma tissue. In a subsequent study, we showed that patients with low metabolic parotid glands, quantified in pre-treatment ^{18}F FDG-PET IBMs, were more likely to develop late xerostomia. These associations also suggested that the non-functional (which can be fatty tissue) to functional tissue ratio is an important pre-treatment characteristic to improve prediction of xerostomia [15].

MRI is superior in imaging soft tissue contrast and therefore more accurate in differentiating fat from the parenchymal gland tissue compared to CT and ^{18}F FDG-PET [16]. Hence, investigating the pre-treatment MR-IBMs of the parotid glands could, therefore, potentially provide better information for predicting late xerostomia.

The purpose of this study was to test whether MR-IBMs extracted from T1-weighted MRI scans were associated with the development of xerostomia 12 months after radiotherapy ($\text{Xer}_{12\text{m}}$) and to investigate whether MR-IBMs can



improve the xerostomia prediction model based on parotid gland dose and baseline xerostomia only. The predictive MR-IBMs were evaluated to test the hypothesis that the fat-to-functional parenchymal parotid tissue ratio is related to Xer_{12m} . The findings were externally validated in an independent cohort.

Materials and methods

Patient demographics and treatment

The training and test cohort included head and neck cancer patients that were treated with definitive radiotherapy with or without concurrent chemotherapy or cetuximab between September 2012 and December 2014 at the University Medical Center Groningen (UMCG), and between October 2010 and March 2016 at Memorial Sloan Kettering Cancer Center (MSKCC), respectively. All patients were treated with Intensity-Modulated Radiation Therapy (IMRT) or Volumetric Arc Therapy (VMAT) using a simultaneous integrated boost (SIB) technique. The parotid glands were spared as much as possible. Patients received a total therapeutic dose of 70 Gy over 6-7 weeks. Most patients received bilateral neck radiation with a prophylactic dose of 54.25 Gy. Details about the radiotherapy regimens used are described in detail in previous studies [14,17].

Patients were excluded if they had salivary gland tumours or underwent surgery or radiotherapy in the head and neck area prior to or within one year after treatment. Moreover, patients without late follow-up data were excluded. Furthermore, MRI scan quality was visually evaluated, and if scans had considerable noise, limiting both visualization of the parotid glands and reliable estimation of the local image intensity, patients were excluded. The final number of patients were 68 and 25 in the UMCG and MSKCC cohorts, respectively.

Endpoints

The primary endpoint was patient-rated moderate-to-severe late xerostomia (Xer_{12m}). In the UMCG cohort, this corresponds to the 2 highest scores of the 4-point Likert scale of the EORTC QLQ-H&N35 questionnaire and was consistently scored 12 months after treatment, which is part of a Standard Follow-up Program (SFP) for Head and Neck Cancer Patients (NCT02435576), as described in previous studies [4,18].

In the MSKCC cohort, xerostomia was scored with multiple questions with a 0-10 scale [19,20] (see supplemental materials 1). Xerostomia scores were

collected between 6 and 17 months after treatment (mean±SD: 11.0±2.5 months). Moderate-to-severe xerostomia was considered if any of the questions was scored 6 or higher.

MRI acquisition and standardization

In the UMCG, MR images were acquired in treatment position on a single scanner (MAGNETOM Aera 1.5T scanner, Siemens Medical Systems, Knoxville, TN, USA) approximately 2 weeks before the start of radiotherapy (Spine 32, flexible 4 and 18 channel coils) for delineation purposes. T1-weighted Turbo Spin Echo (TSE) images (TE: 22ms; TR: 457-606ms) were acquired for all patients with a resolution of 0.36x0.36x4.00 mm without the use of intravenous contrast agents or fat-suppression.

In MSKCC, pre-treatment MR-images were acquired on MRI scanners of different manufacturers (GE, Phillips, Siemens) and scanners with field strength of 1.5T (13 patients) and 3T (12 patients). The resolution of the non-contrast enhanced T1-weighted TSE images (TE: 8-20ms; TR: 400-697ms) ranged from 0.35x0.35 to 1.01 x1.01 mm in-plane and the slice thickness from 3.0 to 5.0 mm.

The MRI intensity values of similar tissue types vary between scans. Therefore, only relative intensities within one scan can be compared. To make a comparison of the relative intensities between patients possible, scans had to be standardized. In this study, fat T1 characteristics were assumed consistent between patients, and should, consequently, have similar MR-intensity values. Subcutaneous fat was delineated in both the right and left cheek area in a minimum of 4 slices at the level of the parotid glands of all patients (Figure 1-I). The fat area was delineated laterally of the parotid gland, the masseter muscle and lip muscles, where the area is delineated as large as possible while excluding non-fat related structures. Subsequently, the MR images were multiplied by a fixed value, which was arbitrarily chosen to 350, and divided by the average subcutaneous fat intensity value. This approach is a simplified tissue-based MRI Intensity standardisation [21]; to our knowledge, no MRI standardisation approaches are known for the head and neck area or radiomics purposes.

Candidate MRI-IBMs, dose and clinical parameters

Parotid glands were delineated for clinical planning purposes on the planning CT, according to guidelines of Brouwer et al. [22]. Dosimetric parameters were extracted from these volumes. MR images were rigidly matched to the CT



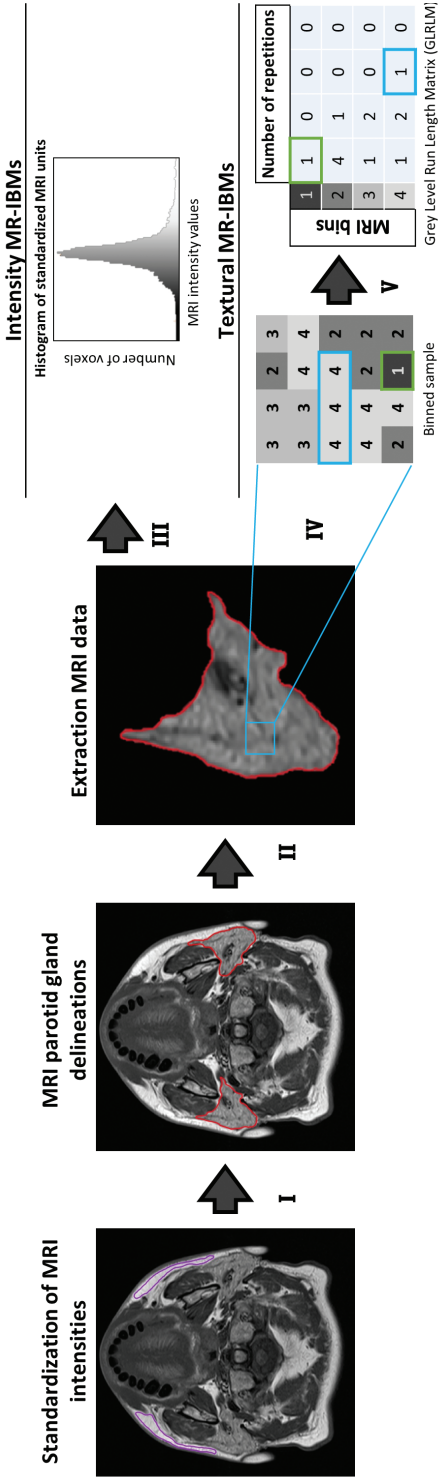


Figure 1 MR-IBM extraction process. I) MR scans were standardized with the average MR intensity, obtained from two delineated fat regions (purple). II) MRI delineated parotid glands were extracted. III) Intensity MR-IBMs were directly extracted IV) The MR intensities were binned. V) Textural MR-IBMs were extracted from matrices reflection grey level transitions or repetitions.

scans, and the CT contours were transferred to the MR images. The MRI parotid delineations were manually corrected where necessary in both datasets.

MRI characteristics of the delineated parotid glands were quantified in intensity and texture MR-image biomarkers (MR-IBMs). Intensity MR-IBMs represent first-order MR-intensity characteristics, such as the mean, minimum, maximum, standard deviation and root mean square of the MR-intensity values.

Furthermore, the MR-intensity heterogeneity was quantified by the textural MR-IBMs. These were extracted from the grey level co-occurrence matrix (GLCM) [23] and grey level run-length matrix (GLRLM) [24,25]. Where, GLCM describes the grey level transitions, GLRLM describes the directional grey level repetitions. The texture IBMs were evaluated in 2D only, which means that the average of MR-IBM values from GLCM and GLRLM of 4 independent directions in-plane were used. Intensity values were discretized from 0 to 450 with a bin size of 25 standardised MR-units [26].

For the complete list of the 21 intensity and 43 textural MR-IBMs (25 GLCM and 18 GLRLM) see supplementary data 2. The extraction process (Figure 1) was performed in MATLAB 2014a and all definitions and formulas were according to the 'Image biomarker standardisation initiative' [27].

Multivariable analysis and model performance

Reference model

A multivariable logistic regression reference model based on the mean dose to the both parotid glands and patient-reported xerostomia at the start of radiotherapy ($Xer_{baseline}$) was fitted in the training cohort [3,4]. $Xer_{baseline}$ was dichotomized as none vs. any in the UMCG dataset and larger than 1 in the MSKCC dataset.

Intensity and textural MR-IBMs selection

To understand the contribution of the different types of MR-IBMs to the reference model, intensity and textural based MR-IBM models were considered separately. Model training was performed in the UMCG cohort only. MR-IBM values were normalised by subtracting each value by the average IBM value and then dividing by standard deviation of that IBM variable.

A pre-selection based on (Pearson) correlation was performed to reduce the effects of overfitting and multicollinearity. If the correlation between two



candidate MR-IBMs was larger than 0.80, only the variable with the highest association with Xer_{12m} was selected.

Multivariable logistic analysis of the pre-selected MR-IBMs was performed together with the mean parotid dose and $Xer_{baseline}$. Based on largest significant log-likelihood differences, step-wise forward selection was used to select MR-IBM predictors [28] (p-value <0.01).

The internal validity of the variable selection was estimated by repeating the entire variable selection procedure (variable normalization, pre-selection and forward selection) 1000 times with a bootstrap procedure with replacement (i.e. with repetition and same sample size). The most frequently selected variables were considered the final model. Model optimism was estimated by calculating the average difference between the performance of the models in each bootstrap and in the original sample, as suggested by the TRIPOD statement [29].

Trained on the UMCG cohort, the MR-IBM models were externally validated in the MSKCC cohort. The model performance measures were the area under the ROC (receiver operating characteristic) curve (AUC), the Nagelkerke's R^2 and the discrimination slope. Model calibration was tested with the average slope and intercept of the models trained on the bootstrap samples that were tested on the original data. The coefficients were corrected for optimism accordingly. In addition, the model improvement was determined with the Likelihood-ratio test, Integrated Discrimination Improvement (IDI) and DeLong's test (testing if AUC significantly improves). The R-packages Regression Modelling Strategies (version 4.3-1) [30] and pROC (version 1.8) were used for these purposes.

Inter-variable relationships

The relation between predictive IBMs and $Xer_{baseline}$ were investigated with the Pearson correlation and univariable logistic analysis, respectively.

Results

Patient characteristics are depicted in Table 1. Generally, all patients received bilateral irradiation. The majority of the patients had oropharyngeal carcinomas and did not report any $Xer_{baseline}$ (59% in the UMCG cohort; 56% in the MSKCC cohort). Moderate-to-severe xerostomia 12 months after radiotherapy (Xer_{12m}) was reported by 34 (50%) of the 68 patients in the UMCG cohort and by 10 (40%) of 25 patients from the MSKCC cohort. In addition, the average (\pm standard

Table 1 Patient characteristics

Characteristics	UMCG		MSKCC	
	N=68	%	N=25	%
Sex				
Female	27	40	5	20
Male	41	60	20	80
Age				
18 - 65 years	47	69	23	92
> 65 years	21	31	2	8
Tumour site				
Oropharynx	42	62	17	68
Nasopharynx	5	7	7	28
Hypopharynx	6	9	-	-
Larynx	10	15	-	-
Oral cavity	2	3	-	-
Other	3	4	1	4
Tumour classification				
Tx	-	-	2	8
T1	11	16	8	32
T2	20	29	8	32
T3	16	24	4	16
T4	21	31	3	12
Node classification				
N0	19	28	7	28
N1	6	9	4	16
N2	42	62	14	56
N3	1	1	-	-
Systemic treatment				
yes	42	62	22	88
no	22	32	3	12
cetuximab	4	6	-	-
Treatment technique				
IMRT	60	88	15	60
VMAT	8	12	10	40
Neck irradiation				
Bilateral	62	91	20	80
Unilateral	1	1	4	16
No	5	7	1	4
Baseline Xerostomia				
No	40	59	14	56
Any	28	41	11	44

Abbreviations: IMRT: Intensity-Modulated Radiation Therapy; VMAT: Volumetric Arc Therapy; UMCG: University Medical Center Groningen; MSKCC: Memorial Sloan Kettering Cancer Center



deviation) mean PG dose was $31.8 \pm 10.9\text{Gy}$ and $22.0 \pm 8.8\text{Gy}$ in the UMCG and MSKCC cohort, respectively. Mean dose to both parotid glands performed slightly better than the contra-lateral gland in this cohort, probably due to the tumour location (oropharynx) and advance N-stage, resulting in comparable contra- and ipsi-lateral doses.

The reference model based on mean PG dose and $\text{Xer}_{\text{baseline}}$ was fitted to the training dataset. The model characteristics and the performance measures ($\text{AUC}=0.81(95\% \text{CI}:0.71-0.91)$, $R^2=0.39$) are depicted in Table 2 and 3. The reference model showed reduced performance in the external dataset ($\text{AUC}_{\text{external.val.}}=0.65(0.41-0.88)$, $R^2_{\text{external.val.}}=0.07$).

Resulting from the bootstrapped variable selection of the intensity MR-IBMs, the 90th intensity percentile (P90) of standardized MRI-units to fat tissue was most frequently selected (175 times of 1000 bootstrapped samples; see Supplementary data 3 for frequency plots). This MR-IBM had both a univariable ($\text{OR}=1.03(1.01-1.05)$; $p=0.004$) and multivariable (Table 2) association with $\text{Xer}_{12\text{m}}$. The positive regression coefficient reveals that high P90 is associated with higher risk of developing $\text{Xer}_{12\text{m}}$ (Table 2). Figure 2 depicts example patients with high and low P90 values.

The P90 added significantly to the variables of the reference model (Likelihood-ratio test; $p=0.002$; IDI; $p=0.004$), and resulted in a substantial and significant improvement of the model performance measures (DeLong's test; $p=0.04$), increasing the reference model's AUC of $0.81(95\% \text{CI}:0.71-0.91)$; $R^2=0.39$, $\text{AUC}_{\text{internal.val.}}=0.78$) to $0.88(0.79-0.96)$; $R^2=0.51$, $\text{AUC}_{\text{internal.val.}}=0.84$) for the intensity MR-IBM model (Table 3). The NTCP-curves for different P90 values are depicted in Figure 3. The performance of the P90 model remained good when externally validated in the MSKCC dataset ($\text{AUC}_{\text{external.val.}}=0.83(0.66-0.99)$, $R^2_{\text{external.val.}}=0.36$). In addition, univariable analysis in the external dataset showed a significant association of P90 with $\text{Xer}_{12\text{m}}$ ($p=0.039$).

From the texture MR-IBMs, the Grey Level Non-uniformity Normalised (GLN_{nor}) was most often selected (91 of 1000 bootstrapped samples; see Supplementary data 3). Derived from the GLRLM, this texture MR-IBM is high when high concentrations of runs with the same grey level are present in the volume of interest (for formula see Supplementary data 2). This texture MR-IBM had both a negatively significant univariable ($\text{OR}=0.34$, $95\% \text{CI } 0.20-0.74$; $p=0.004$) and

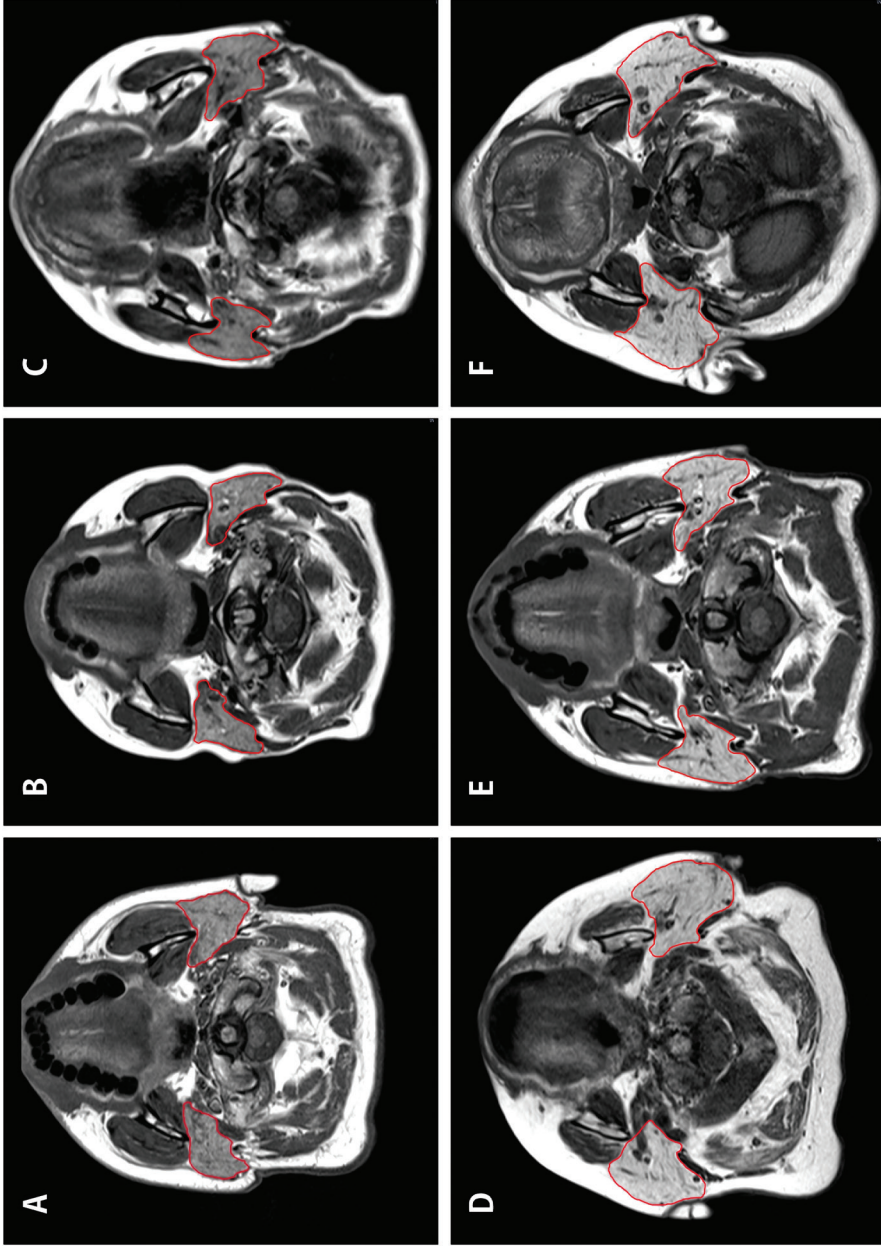


Figure 2 Examples of patients with low (A-C) and high (D-F) P90 values of the parotid glands. Accordingly, the patients in the bottom row are more at risk of developing late xerostomia than those in the top row.



multivariable (Table 2) association with Xer_{12m} , indicating that low GLN_{nor} values were related with a higher risk of xerostomia.

When adding GLN_{nor} to the reference model, model performance significantly improved (Likelihood-ratio test; $p=0.002$; IDI; $p<0.001$). The performance of the resulting texture MR-IBM model (Table 2) was good (AUC: 0.88(0.79-0.96), $R^2=0.52$; $AUC_{internal.val.}=0.84$) and significantly improved compared to the reference model (DeLong's test; $p=0.03$). The NTCP curves are depicted in Figure 3.

On external validation, the texture MR-IBM model performed well ($AUC_{external.val.}=0.83(0.67-0.99)$, $R^2_{external.val.}=0.31$). Univariable analysis also showed a significant association of GLN_{nor} with Xer_{12m} ($p=0.036$).

The internal validation calibration slope and intercept showed reasonable goodness-of-fit for both the intensity and texture models (Table 3). In addition, all models were fitted to the combined dataset (MSKCC + UMCG), and showed similar coefficients and performance measures (Supplementary data 4).

The intensity and texture MR-IBM P90 and GLN_{nor} were highly correlated ($r=-0.85$ (95%CI: -0.90 – -0.78); $p<0.001$). In a multivariable analysis, the addition of GLN_{nor} did not add significant information in predicting Xer_{12m} with P90 and vice versa (Likelihood-ratio test; $p>0.27$). Furthermore, univariable logistic analysis showed no significant association between $Xer_{baseline}$ and P90 ($p=0.45$) or GLN_{nor} ($p=0.29$).

Table 2 Model characteristics of prediction models with and without MR-IBMs trained in UMCG cohort

	Reference model			Intensity MR-IBM model			Texture MR-IBM model		
	β Apparent	OR (95% CI)	p-value	β Apparent	OR (95% CI)	p-value	β Apparent	OR (95% CI)	p-value
intercept	-3.311	-2.859		-11.30	-8.540		-3.642	-2.673	
Xer _{baseline}	2.351	10.50 (3.06-36.06)	<0.001	2.734	2.063	15.39 (3.56-66.47)	2.736	2.009	15.42 (3.44-69.07)
PG dose	0.074	1.08 (1.01-1.14)	0.015	0.072	0.054	1.07 (1.00-1.15)	0.080	0.059	1.08 (1.01-1.16)
P90	-	-	-	0.034	0.026	1.03 (1.01-1.06)	-	-	-
GLN _{hor}	-	-	-	-	-	-	-1.146	-0.841	0.32 (0.14-0.71)

Corr. corrected for optimism with bootstrapping; MR-IBM: Magnetic Resonance Image Biomarker; Xerbaseline: xerostomia at baseline; PG dose: mean dose to parotid glands; P90: 90th percentile of MR intensities; GLNhor: Grey Level Non-uniformity Normalised; β : regression coefficients; OR: Odds Ratio; CI: Confidence Interval



Table 3 Performance of prediction models with and without MR-IBMs in training (UMCG) and external validation cohort (MSKCC)

	Reference model			Intensity MR-IBM model			Texture MR-IBM model		
	Xerbaseline + PG dose			Xerbaseline + PG dose + P90			Xerbaseline + PG dose + GLNnor		
	UMCG	MSKCC		UMCG	MSKCC		UMCG	MSKCC	
	Apparent	Int. val.	Ext. val.	Apparent	Int. val.	Ext. val.	Apparent	Int. val.	Ext. val.
Area Under the Curve (AUC)	0.81 (0.71-0.91)	0.78	0.65 (0.41-0.88)	0.88 (0.79-0.96)	0.83	0.83 (0.66-0.99)	0.88 (0.79-0.96)	0.84	0.83 (0.67-0.99)
Nagelkerke R ²	0.39 (0.21-0.39)	0.32	0.07	0.51 (0.41-0.61)	0.39	0.36	0.52 (0.42-0.61)	0.39	0.31
Discrimination slope	0.31	0.27	0.11	0.42	0.35	0.27	0.43	0.37	0.24
Calibration slope (intercept)	-	0.87 (0.00)	0.45 (-0.18)	-	0.75 (0.01)	1.11 (-0.93)	-	0.73 (0.00)	0.94 (-0.28)
LLH-ratio test	-	-	-	9.34 (p=0.002)	-	6.45 (p=0.011)	9.97 (p=0.002)	-	5.05 (p=0.025)
IDI	-	-	-	0.11 (p=0.004)	-	0.16 (p=0.035)	0.12 (p<0.001)	-	0.14 (p=0.025)
DeLong's test	-	-	-	-1.72 (p=0.043)	-	-1.87 (p=0.030)	-1.87 (p=0.030)	-	-2.09 (p=0.018)

MR-IBM: Magnetic Resonance Image Biomarker; Xerbaseline: xerostomia at baseline; PG dose: mean dose to parotid glands; P90: 90th percentile of MR intensities; GLNnor: Grey Level Non-uniformity Normalised; UMCG: University Medical Center Groningen; MSKCC: Memorial Sloan Kettering Cancer Center; Internal val.: corrected for optimism with bootstrapping; int. and Ext. val.: internally and externally validated; LLH: likelihood

Discussion

In previous studies we showed that more heterogeneous CT intensity characteristics and low metabolic ^{18}F FDG-PET activity of the parotid glands were related to a higher risk of xerostomia 12 months after radiotherapy ($\text{Xer}_{12\text{m}}$) [14,15]. These findings led to the hypothesis that the fat-to-functional parenchymal parotid tissue ratio is an important pre-treatment marker to improve prediction of $\text{Xer}_{12\text{m}}$. The results of the current study also support this hypothesis.

Other recent studies also showed that pre-treatment information extracted from CT images, quantifying the parotid gland texture [31] and shape [32], were associated with observed radiation-induced xerostomia. Additionally, several studies have shown associations between xerostomia and parotid gland changes in CT data [33,34]. The current study is novel by investigating pre-treatment MR intensities of the parotid glands, providing high contrast soft-tissue information, in relation to late patient-rated xerostomia.

MRI characteristics of the parotid glands, quantified in pre-treatment MR-IBMs, were significantly associated to $\text{Xer}_{12\text{m}}$. Moreover, the $\text{Xer}_{12\text{m}}$ prediction improved with the addition of MR-IBMs to the reference model using mean parotid glands dose and baseline complaints only (from an AUC of 0.81 to 0.88). These results were also valid in an independent external cohort, where the performance of the reference model ($\text{AUC}_{\text{external.val.}}=0.65$) was low compared to the MR-IBM models ($\text{AUC}_{\text{external.val.}}=0.83$). This underlines the importance of tissue-specific characteristics in predicting and understanding the development of radiation-induced toxicities, which is becoming increasingly important in the selection of patients for more advanced radiation techniques [6,7] and to tailor the treatment to the patient specifically [5].

The most frequently selected intensity MR-IBM was the P90, indicating the 90th percentile of the MR-intensities of the parotid glands. Since the MR-intensity values were standardized to fat, fat tissue can be assumed to have comparable MR-intensity between patients. Since fat has a short T1 relaxation time compared to parenchymal or muscle tissue, it is presented with a high signal intensity in T1-weighted images [35]. Hence high P90 values relate to high fat concentration in the parotid gland. More specifically, if at least 10% of the volume of the parotid gland has high intensity values, patients were at higher risk of developing late xerostomia. However, this volume percentage should be evaluated with caution, since using the simpler 'mean standardized T1 intensity'



also significantly improved the reference model (AUC=0.86; Likelihood-ratio test: $p=0.005$). Moreover, 16 of the 21 intensity MR-IBMs and 34 of the 43 texture MR-IBMs also contributed significantly, as single variables, to the reference model in predicting Xer_{12m} (supplementary data 5). This indicates that other MR-IBMs that are also related to parotid gland intensity and texture, can give similar results as P90 and GLN_{nor} .

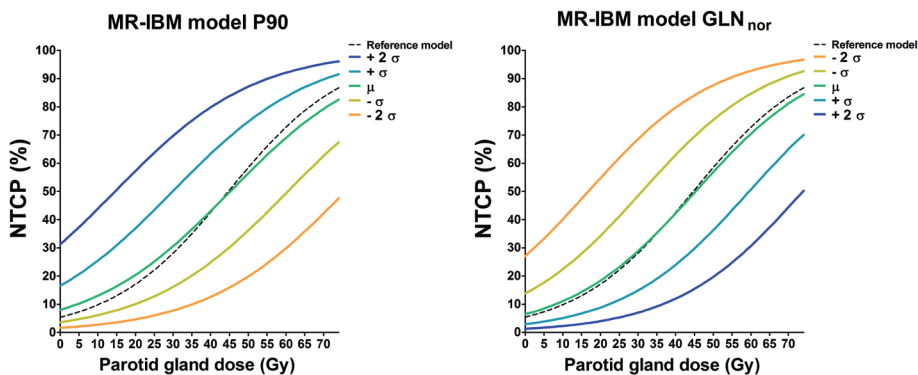


Figure 3 Normal Tissue Complication Probability (NTCP) models for Xer_{12m} based on the MR-IBM P90 (left) and GLN_{nor} (right) (Table 2). Plotted against the mean dose of both parotid gland, NTCP curves are given for the mean (in green) plus/minus one (turquoise/yellow) and two (blue/orange) standard deviations of the P90 ($\mu=234.89$, $\sigma=31.80$) and GLN_{nor} ($\mu=0.18$, $\sigma=0.03$) values for $Xer_{baseline}=0$.

The values of the selected texture MR-IBM, GLN_{nor} are low if grey values are equally distributed over all grey levels, i.e. more heterogeneity. Lower GLN_{nor} values were associated with a higher risk of developing xerostomia. This was also demonstrated in a previous study based on CT parotid gland IBMs [14]. Noteworthy, the reverse of GLN_{nor} is ‘entropy’, which was the second most frequently selected intensity MR-IBM (Supplementary data 3). Moreover, GLN_{nor} was highly correlated with P90, which suggests that parotid glands with high (fat related) MR signal intensities were more heterogeneous. More research in larger datasets is necessary to determine whether both characteristics are relevant in the development of xerostomia and the generalizability of P90 and GLN_{nor} , or whether they reflect similar biological information.

The theoretical and qualitative evaluation of the predictive MR-IBMs suggested a relation with the fat concentration and heterogeneity of the parotid gland. In previous work, unrelated to the oncology field, Izumi et al. [36] presented an MRI-

based grading of the severity of parotid impairment for patients with Sjögren's syndrome that was based on similar image characteristics: high T1-weighted signal intensity areas (e.g. fat tissue) and heterogeneity in the parotid glands. Another study by Izumi et al. [37] also showed a relationship between increased signal intensities on T1-weighted MR images and impaired parotid function for patients suffering from hyperlipidaemia. The findings of the current and studies suggest that increased fat concentration in the parotid gland, which may be caused by parenchymal changes due to lipid infiltration, can increase the probability of developing xerostomia after radiotherapy.

MRI offers the advantage of non-invasively acquired images with high soft tissue contrast without the use of radiation with respect to CT and PET imaging. However, it is a complex image modality due to the large range of possible acquisition settings, and requires intensity standardisation. This study and previous studies have both indicated that PET and MR IBMs seem to perform better than the CT IBMs in identifying patients that develop xerostomia [14,15]. Studies including IBMs from all three image modalities are necessary to determine which modality is most optimal in this context, or whether they can add to each other in predicting late xerostomia. The analyses of the current study were based on relatively simple T1-weighted TSE, which is widely used and requires no administration of intravenous contrast agents. However, more sophisticated MRI sequences may better differentiate between fat and functional parotid tissue (i.e. combinations of non- and fat saturated images or functional information (e.g. DIXON, Diffusion Weighted or Dynamic contrast-enhanced imaging)). In addition, IBMs extracted from wavelet transformed images might improve the performance of the models presented in the current study.

Limitations of the present study are the small cohort sizes and the lack of one-to-one correspondence in xerostomia assessments between the two cohorts. However, a careful matching was performed such that the two moderate-to-severe assessments would be as similar as possible given the data. An additional limitation is the large variability in MR acquisition parameters in the MSKCC cohort compared to the training cohort. Firstly, the resolution had a relative large range in these scans. This can impact the texture IBMs, which depend on the spatial intensity distribution [38,39]. Secondly, patients in the MSKCC cohort were scanned without a thermoplastic mask, resulting in parotid glands deformation due to the music headphones that patients wore during acquisition. Finally, MSKCC scans were acquired with different field strengths, and scanners



from different vendors. Even though part of this variability should be captured by MRI standardisation, this can influence the MRI intensity and contrast. Despite these limitations, the performance of the MR-IBM models was good when tested in the MSKCC dataset, suggesting that these IBMs were robust to variability in image acquisition parameters. The simplicity of the P90 metric likely contributed to successful validation.

Driven by the hypothesis, the MRI intensity standardisation was linearly performed to ensure similar fat tissue intensities between patients. However, this is in reality not a linear problem [40]. Our approach is simple, and could be regarded as a starting point to improve the standardisation so that not only subcutaneous fat is generalized between patients, but also other tissues, such as muscle. Additionally, mainly due to the presence of field inhomogeneity's, scans can have intensity variations within the scan, for which sophisticated bias field correction algorithms have been developed for brain MR images [41]. The above described corrections were explored for this dataset, however, the effect of these corrections on IBM analysis is currently unknown, and needs further investigation.

In conclusion, the results of the current study support the hypothesis that a high fat concentration, quantified in MR-IBMs, within the parotid glands is related to a higher risk of developing xerostomia 12 months after radiotherapy (Xer_{12m}). The prediction performance of Xer_{12m} based on parotid dose and baseline xerostomia only was improved by the addition of the predictive intensity MR-IBM P90. These results were maintained in a small external validation cohort. MR-IBMs appear to be good candidates to predict the patient-specific response of healthy tissue to radiation dose. However, more research in larger patient cohorts is needed to further validate our conclusions.

References

- 1 Dirix P, Nuyts S. Evidence-based organ-sparing radiotherapy in head and neck cancer. *Lancet Oncol* 2010;11:85–91.
- 2 Hawkins PG, Lee JY, Mao Y, Li P, Green M, Worden FP, 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. *Radiother Oncol* 2017;126:68–74.
- 3 Houweling AC, Philippens MEP, Dijkema T, Roesink JM, Terhaard CHJ, Schilstra C, et al. A comparison of dose-response models for the parotid gland in a large group of head-and-neck cancer patients. *Int J Radiat Oncol Biol Phys* 2010;76:1259–65.
- 4 Beetz I, Schilstra C, Van Der Schaaf A, Van Den Heuvel ER, Doornaert P, Van Luijk P, 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. *Radiother Oncol* 2012;105:101–6.
- 5 Langendijk JA, Lambin P, De Ruyscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. *Radiother Oncol* 2013;107:267–73.
- 6 Lomax A. Intensity modulation methods for proton radiotherapy. *Phys Med Biol* 1999;44:185–205.
- 7 Legendijk JJW, Raaijmakers BW, Raaijmakers AJE, Overweg J, Brown KJ, Kerkhof EM, et al. MRI/linac integration. *Radiother Oncol* 2008;86:25–9.
- 8 Abgral R, Keromnes N, Robin P, Le Roux P-Y, Bourhis D, Palard X, et al. Prognostic value of volumetric parameters measured by (18)F-FDG PET/CT in patients with head and neck squamous cell carcinoma. *Eur J Nucl Med Mol Imaging* 2014;41:659–67.
- 9 Jeong J, Setton JS, Lee NY, Oh JH, Deasy JO. Estimate of the impact of FDG-avidity on the dose required for head and neck radiotherapy local control. *Radiother Oncol* 2014;111:340–7.
- 10 Koyasu S, Nakamoto Y, Kikuchi M, Suzuki K, Hayashida K, Itoh K, et al. Prognostic value of pretreatment 18F-FDG PET/CT parameters including visual evaluation in patients with head and neck squamous cell carcinoma. *AJR Am J Roentgenol* 2014;202:851–8.
- 11 Alluri KC, Tahari AK, Wahl RL, Koch W, Chung CH, Subramaniam RM. Prognostic value of FDG PET metabolic tumor volume in human papillomavirus-positive stage III and IV oropharyngeal squamous cell carcinoma. *AJR Am J Roentgenol* 2014;203:897–903.
- 12 Zhai T-T, van Dijk L V, Huang B-T, Lin Z-X, Ribeiro CO, Brouwer CL, et al. Improving the prediction of overall survival for head and neck cancer patients using image biomarkers in combination with clinical parameters. *Radiother Oncol* 2017;256–62.
- 13 Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5.



Chapter 4 - Parotid gland fat related MR-IBMs improve prediction of late xerostomia

- 14 van Dijk L V., Brouwer CL, van der Schaaf A, Burgerhof JGM, Beukinga RJ, Langendijk JA, et al. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. *Radiother Oncol* 2017;122:185–91.
- 15 van Dijk L V., Noordzij W, Brouwer CL, Boellaard R, Burgerhof JGM, Langendijk JA, et al. 18F-FDG PET image biomarkers improve prediction of late radiation-induced xerostomia. *Radiother Oncol* 2017;126:89–95.
- 16 Burke CJ, Thomas RH, Howlett D. Imaging the major salivary glands. *Br J Oral Maxillofac Surg* 2011;49:261–9.
- 17 Van Der Laan HP, Gawryszuk A, Christianen MEMC, Steenbakkers RJHM, Korevaar EW, Chouvalova O, et al. Swallowing-sparing intensity-modulated radiotherapy for head and neck cancer patients: Treatment planning optimization and clinical introduction. *Radiother Oncol* 2013;107:282–7.
- 18 Vergeer MR, Doornaert PAH, Rietveld DHF, Leemans CR, Slotman BJ, Langendijk JA. Intensity-modulated radiotherapy reduces radiation-induced morbidity and improves health-related quality of life: results of a nonrandomized prospective study using a standardized follow-up program. *Int J Radiat Oncol Biol Phys* 2009;74:1–8.
- 19 Tam M, Riaz N, Kannarunimit D, Pena AP, Schupak KD, Gelblum DY, et al. Sparing bilateral neck level IB in oropharyngeal carcinoma and xerostomia outcomes. *Am J Clin Oncol* 2015;38:343–7.
- 20 Eisbruch A, Kim HM, Terrell JE, Marsh LH, Dawson LA, Ship JA. Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2001;50:695–704.
- 21 Robitaille N, Mouiha A, Crépeault B, Valdivia F, Duchesne S. Tissue-based MRI intensity standardization: Application to multicentric datasets. *Int J Biomed Imaging* 2012;2012.
- 22 Brouwer CL, Steenbakkers RJHM, Bourhis J, Budach W, Grau C, Grégoire V, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother Oncol* 2015;117:83–90.
- 23 Haralick R, Shanmugan K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973;3:610–21.
- 24 Tang X. Texture information in run-length matrices. *IEEE Trans Image Process* 1998;7:1602–9.
- 25 Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process* 1975;4:172–9.
- 26 Leijenaar RTH, Nalbantov G, Carvalho S, van Elmpt WJC, Troost EGC, Boellaard R, et al. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. *Sci Rep* 2015;5:11075.
- 27 Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative - feature definitions. [arXiv:161207003](https://arxiv.org/abs/1612.07003) 2016.

- 28 Dehing-Oberije C, De Ruyscher D, Petit S, Van Meerbeeck J, Vandecasteele K, De Neve W, et al. Development, external validation and clinical usefulness of a practical prediction model for radiation-induced dysphagia in lung cancer patients. *Radiother Oncol* 2010;97:455–61.
- 29 Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* 2015;162:W1.
- 30 R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: the R Foundation for Statistical Computing. 2011:Available online at <http://www.R-project.org/>.
- 31 Nardone V, Tini P, Nioche C, Mazzei MA, Carfagno T, Battaglia G, et al. Texture analysis as a predictor of radiation-induced xerostomia in head and neck patients undergoing IMRT. *Radiol Medica* 2018:1–9.
- 32 Gabryś HS, Buettner F, Sterzing F, Hauswald H, Bangert M. Design and Selection of Machine Learning Methods Using Radiomics and Dosiomics for Normal Tissue Complication Probability Modeling of Xerostomia. *Front Oncol* 2018;8:1–20.
- 33 Scalco E, Fiorino C, Cattaneo GM, Sanguineti G, Rizzo G. Texture analysis for the assessment of structural changes in parotid glands induced by radiotherapy. *Radiother Oncol* 2013;109:384–7.
- 34 Broggi S, Fiorino C, Dell’Oca I, Dinapoli N, Paiusco M, Muraglia A, et al. A two-variable linear model of parotid shrinkage during IMRT for head and neck cancer. *Radiother Oncol* 2010;94:206–12.
- 35 [Http://www.startradiology.com/the-basics/mri-technique/](http://www.startradiology.com/the-basics/mri-technique/). MRI Technique n.d.
- 36 Izumi M, Eguchi K, Ohki M, Uetani M, Hayashi K, Kita M, et al. MR imaging of the parotid gland in Sjögren’s syndrome: A proposal for new diagnostic criteria. *Am J Roentgenol* 1996;166.
- 37 Izumi M, Hida A, Takagi Y, Kawabe Y, Eguchi K, Nakamura T. MR imaging of the salivary glands in sicca syndrome: comparison of lipid profiles and imaging in patients with hyperlipidemia and patients with Sjogren’s syndrome. *AJR AmJRoentgenol* 2000;175:829–34.
- 38 Shafiq-UI-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, et al. Intrinsic dependencies of CT radiomics features on voxel size and number of gray levels Supplementary Material : *Med Phys* 2017:1–10.
- 39 Mackin D, Fave X, Zhang L, Fried D, Taylor B, Rodriguez-rivera E, et al. Measuring CT scanner variability of radiomics features. *Invest Radiol* 2015;50:757–65.
- 40 Nyúl LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. *IEEE Trans Med Imaging* 2000;19:143–50.
- 41 Li C, Gore JC, Davatzikos C. Multiplicative intrinsic component optimization (MICO) for MRI bias field estimation and tissue segmentation. *Magn Reson Imaging* 2014;32:913–23.



Chapter 5

Geometric image biomarker changes of the parotid gland are associated with late xerostomia

Published in: **International Journal of Radiation Oncology biology physics**
2017 Augustus; 1101–10

van Dijk LV, Brouwer CL, van der Laan HP, Burgerhof JGM, Langendijk JA, Steenbakkens RJHM, Sijtsema NM.

Online Supplemental Materials: <http://dx.doi.org/10.1016/j.ijrobp.2017.08.003>

Abstract

Background and purpose

The aim of this study was to identify a surrogate marker for late xerostomia 12 months after radiotherapy (Xer_{12m}) based on information obtained shortly after treatment.

Materials and Methods

Differences in parotid gland (PG) were quantified in image biomarkers (Δ IBMs) before and 6 weeks after radiotherapy of 107 patients. By performing step-wise forward selection, Δ IBMs that were associated with Xer_{12m} were selected. Subsequently, other variables, such as PG dose and acute xerostomia scores were added to improve the prediction performance. All models were internally validated.

Results

Prediction of Xer_{12m} based on PG surface reduction (Δ PG-surface) was good (AUC=0.82). PG dose was related to Δ PG-surface ($p < 0.001$, $R^2 = 0.27$). The addition of acute xerostomia scores to the Δ PG-surface improved the prediction of Xer_{12m} significantly and vice versa. The final model including Δ PG-surface and acute xerostomia had outstanding performance in predicting Xer_{12m} early after radiotherapy (AUC=0.90).

Conclusion

PG surface reduction was associated with late xerostomia. The early post-treatment model with Δ PG-surface and acute xerostomia scores can be considered as surrogate marker for late xerostomia.

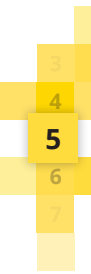
Introduction

Xerostomia is one of the most frequent side effect that affects many head and neck cancer (HNC) patients after radiotherapy and has a major impact on quality of life (1). Limiting the dose to the parotid glands (PG) reduces the probability of developing xerostomia (2–4). Although multiple studies have investigated the relation between dose and the risk of xerostomia, substantial variability in this relationship remains unexplained (2, 3). A possible reason for this variation is that dosimetric parameters (and baseline xerostomia scores) are not the only explaining variables, but that patient-specific characteristics, such as intrinsic radiosensitivity, also affect the development of late xerostomia (5). Unexplained variability could, moreover, result from inconsistency in the assessment of xerostomia, i.e. patient-rated xerostomia, as this is a subjective measure (6). More specifically, the individual experience of a side effect with similar function loss varies widely among individual patients, depending on many aspects, such as interpretation of the questions and general quality of life (7). A more quantitative measure of late xerostomia may lead to improvement of prediction models by increasing the consistency of the endpoint.

A surrogate endpoint early after treatment to evaluate late xerostomia is not only interesting in order to understand the development of xerostomia better, but would also be desirable and beneficial in order to potentially improve the time and cost effectiveness of future clinical studies in HNC patients. Additionally, this could also contribute to the physician–patient dialog at the end of treatment to provide patients with a more reliable prognosis regarding the expected severity of xerostomia for the next few years. Moreover, it can support selection of patients that do not recover from acute xerostomia for potential future therapeutic strategies of xerostomia, such as adult stem cell-based therapy (8, 9).

CT image acquisition, which is routinely used for radiotherapy treatment planning and response assessment, would be an ideal modality to quantify changes of radiated tissues, as it is rapid, relatively cheap and widely available.

To identify quantitative candidate surrogates for assessing xerostomia, PG characteristics were quantified by extracting image biomarkers (IBMs) of the PGs before and after radiotherapy and by calculating the differences (Δ IBMs). The main objective of this study was to test the hypothesis that Δ IBMs – either combined with other predictive factors or not – were associated with late



xerostomia and to test whether an early post-treatment model based on these ΔIBMs could serve as a surrogate marker for late xerostomia.

Materials and methods

Patients

The 107 HNC patients that were prospectively included in this study were treated with radiotherapy, either in combination with concurrent chemotherapy or cetuximab or not, between June 2008 and April 2012. Patients with salivary glands tumours, those that previously (or one year after) underwent surgery or radiotherapy in the head and neck area were excluded from this study. Moreover, patients without follow-up data 12 months after RT were excluded. For a detailed description of the radiation protocols we refer to the paper of Christianen et al (10). Briefly, most patients were treated with IMRT that was optimised to spare the parotid glands without compromising the dose to the target volumes (11, 12), using a simultaneous integrated boost (SIB) technique. Generally, 70 Gy (2 Gy per fraction) was administered to the primary tumour and pathologic lymph nodes over the course of 6 or 7 weeks (6 or 5 fractions per week, respectively). The majority of patients received elective radiation to the cervical lymph node levels of 54.25 Gy (1.55 Gy per fraction) (13). More patient characteristics are depicted in Table 1.

For all patients, a standardized planning CT scan (Somatom Sensation Open, Siemens, Forchheim, Germany; voxel size: 0.94 x 0.94 x 2.0 mm³; 100–140 kV) was acquired 2 weeks before treatment. Six weeks after radiotherapy, a second CT scan was acquired together with the last assessment of acute toxicity. Both scans were acquired in with a thermoplastic mask in radiotherapy treatment positioning. This study was approved the Medical Ethics Commission and all participating patients gave informed consent.

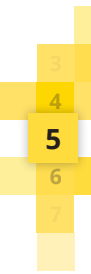
Endpoints

Patient-rated xerostomia scores were evaluated prospectively on a routine basis, before radiotherapy, weekly during radiotherapy, and subsequently 6 weeks and 6 and 12 months after radiotherapy using the EORTC QLQ-H&N35 questionnaire, as part of the standard follow-up programme (SFP) at University Medical Center Groningen (2, 14). The primary endpoint of this study was moderate-to-severe patient-rated xerostomia at 12 months after radiotherapy (Xer_{12m}). This

Table 1 Patient characteristics

Characteristics	N=107	%
Sex		
female	12	11
male	95	89
Age		
18 – 65 years	67	63
> 65 years	40	37
Tumour site		
Oropharynx	28	26
Nasopharynx	3	3
Hypopharynx	7	7
Larynx	65	61
Oral cavity	2	2
Unknown primary	2	2
Tumour classification		
T _{in situ}	1	1
T0	2	2
T1	25	23
T2	47	44
T3	18	17
T4	14	13
Node classification		
N0	67	63
N1	8	7
N2abc	29	27
N3	3	3
Systemic treatment		
yes	28	26
no	79	74
Treatment technique		
3D-CRT	22	21
ST-IMRT	46	43
SW-IMRT	39	36
Neck irradiation		
Bilateral	67	63
Unilateral	6	6
No	34	32

Abbreviations: CRT: Conformal Radiation Therapy; IMRT: Intensity-Modulated Radiation Therapy; ST-IMRT: standard parotid sparing IMRT; SW-IMRT: swallowing sparing IMRT; SW-VMAT: swallowing sparing Volumetric Arc Therapy



corresponds to the 2 highest scores of the 4-point Likert scale (not, a bit, quite a bit, a lot).

Quantification of PG changes in Δ IBMs

The PGs were delineated on the planning CT according to the consensus guidelines of Brouwer et al. (15). Using deformable image registration, delineations were warped to the repeat CT in Mirada RTx (Mirada Medical Ltd., Oxford, UK). The warped contours were manually corrected if necessary.

All image biomarkers were extracted from the planning and the repeat CT with in-house developed software that was implemented in Matlab (version R2014a). Subtraction of the pre- from post-treatment IBMs resulted in the Δ image biomarkers (Δ IBMs). Twenty geometric IBMs of the PGs, such as volume and compactness, were extracted from the delineations directly. Additionally, twenty CT-intensity Δ IBMs, were extracted from the CT data of the PGs. For a list of the Δ IBMs and the additional 8 clinical variables refer to the Supplementary material 1.

Reference model

A reference prediction model for late xerostomia based on the predictors found by Beetz et al. (2) (mean dose to the contralateral PG and the baseline xerostomia) was fitted to the dataset (Figure 1-‘reference model’). The patient-reported xerostomia at start of radiotherapy ($Xer_{baseline}$) was dichotomized as none vs. any. The PG that received the least amount of mean dose was considered contralateral.

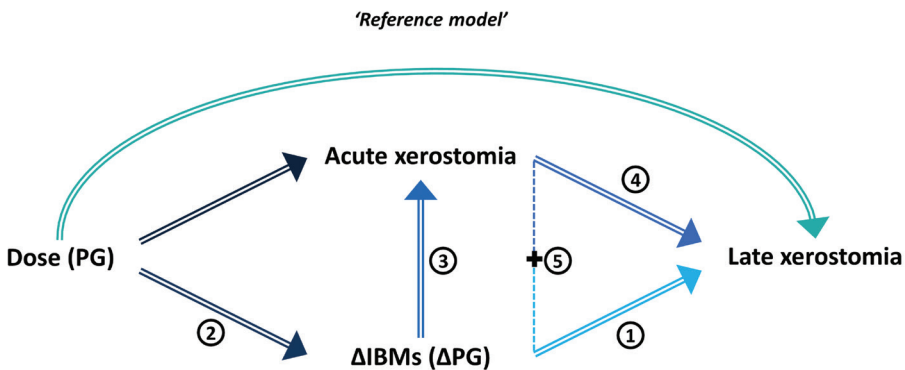


Figure 1 Investigated associations: ① Δ IBMs and late xerostomia (Xer_{12m}), ② PG dose and Δ IBMs, ③ Δ IBMs and acute xerostomia ($Xer_{6w-post}$), ④ acute xerostomia ($Xer_{6w-post}$) and late xerostomia ⑤ Δ IBMs and acute xerostomia combined and late xerostomia, ‘Reference model’ PG dose and late xerostomia. All associations were corrected for baseline xerostomia. (Abbreviations — IBM: image biomarker; PG: parotid gland)

Δ IBM selection for late xerostomia

To investigate the associations between the potential Δ IBMs and Xer_{12m} (Figure 1-①), Δ IBMs were considered as candidate variables in the variable selection process as described below. Subsequently, the modeling process was repeated by adding $Xer_{baseline}$ first and subsequently the mean PG dose to the candidate variables. All individual patient variable values were normalised by subtracting each value by the sample mean and then dividing by the sample standard deviation of that IBM variable.

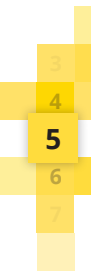
Introducing large numbers of highly correlated variables may have negative effects on variable selection, due to overfitting and multicollinearity (16, 17). Candidate Δ IBMs were therefore pre-selected based on their (Pearson) correlation. If the correlation of two variables was larger than 0.80, only the Δ IBMs with the highest association with Xer_{12m} was selected for further analysis. Step-wise forward selection, based on log-likelihood (18), was used to select the most important predictors (p-value <0.01). The internal validity of the variable selection was estimated with a bootstrap procedure. The entire variable selection procedure (variable normalization, pre-selection and forward selection) was repeated in 1000 bootstrapped samples (i.e. with replacement). From the resulting models, the most frequently selected variables were considered for the final models.

The selected model's optimism was estimated by calculating the difference between the performance of the models in each bootstrap and in the original sample, as suggested by the TRIPOD statement (19).

The model's performance was quantified in terms of discrimination with the area under the ROC (receiver operating characteristic) curve (AUC), the Nagelkerke R^2 and the discrimination slope. Model calibration was tested with the Hosmer-Lemeshow test and by calculating the slope and intercept of a logistic regression model of the linear predictor derived from the predicted probability of moderate-to-severe late xerostomia (variable) against the actual xerostomia outcome (response). The coefficients were corrected for optimism accordingly. The R-packages Regression Modeling Strategies (version 4.3-1) (20) were used for these purposes.

Relation parotid gland dose and selected Δ IBMs

Linear regression was performed to investigate the relation of mean PG dose to the selected Δ IBMs (Figure 1-②). Both PGs were considered separately in



investigating this relation. Model performance was measured as the explained variance (R^2) and normality of the residuals of the regression models was checked.

Δ IBMs and other predictive variables

Firstly, the addition of mean PG dose to the model with selected Δ IBMs and $Xer_{baseline}$ was investigated.

Secondly, the relation between Δ IBMs and acute xerostomia 6 weeks after radiotherapy ($Xer_{6w-post}$, moderate-to-severe) was investigated (Figure 1-③), in order to analyze whether the selected Δ IBMs were a direct substitute measure of acute xerostomia scores. If the assumption that acute and late xerostomia scores are related would be correct (Figure 1-④), then the selected Δ IBMs could actually be a measure of acute xerostomia rather than late xerostomia. Therefore, the presumed assumption was tested by investigating the logistic relation between $Xer_{6w-post}$ and Xer_{12m} . Subsequently, a multivariable analysis and variable selection was performed to investigate whether the $Xer_{6w-post}$ contained additional information to the Δ IBMs to predict late xerostomia (Figure 1-⑤). Baseline xerostomia was also considered for these analyses.

Actual xerostomia incidences were depicted over time to illustrate the effects of the Δ IBMs and the final post-treatment model with the best prediction performance. Patients were classified based on their Δ IBM values (higher or lower than median) and on their predicted risk calculated with the final prediction model (higher or lower than 50%).

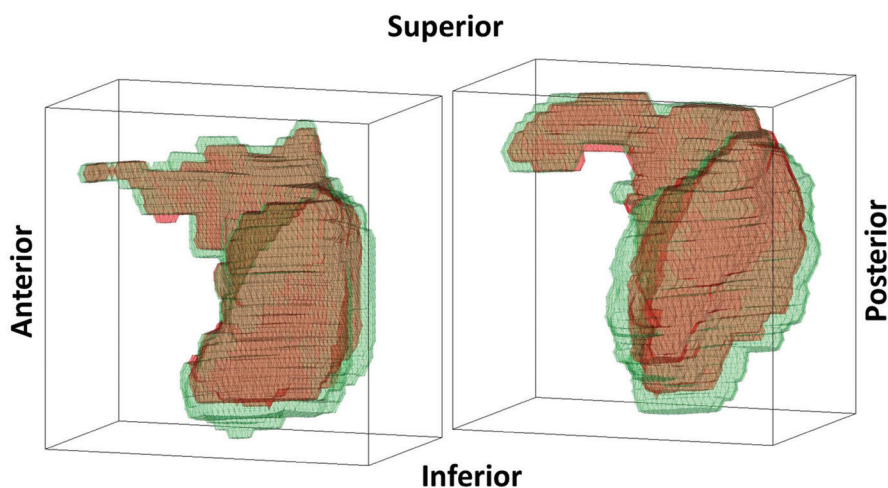


Figure 2 Examples of parotid glands (PG) with a large negative Δ PG-surface (and Δ PG-volume) between the start of radiotherapy (green) and 6 weeks after radiotherapy (red).

Results

Reference model

Moderate-to-severe xerostomia 12 months after radiotherapy (Xer_{12m}) was reported by 32 (30%) of the 107 patients. The reference model based on mean PG dose and $Xer_{baseline}$ was fitted to the dataset. The model characteristics and the performance measures (AUC=0.76, $R^2=0.28$) are depicted in Table 2 (Reference model).

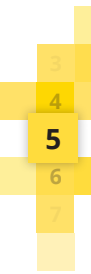
Δ IBMs selection for late xerostomia

The most frequently selected Δ IBM variable was Δ PG-surface (a visual representation is depicted in Figure 2) of the contralateral PG, obtained with forward step-wise selection (468 times of 1000 bootstrapped samples; see supplementary material 2 for frequency plots). This variable was significantly associated to Xer_{12m} (Table 2, $p<0.001$; OR: 0.86 (0.79–0.92)). The regression coefficient was negative, indicating that larger PG surface reduction or shrinkage relates to a higher risk of developing Xer_{12m} . $Xer_{baseline}$ showed a significant predictive contribution in addition to Δ PG-surface (likelihood ratio test; $p<0.001$). A model with Δ PG-surface and $Xer_{baseline}$ was created with good performance (Table 2, AUC=0.82 (0.72 – 0.91), also after internal validation, which includes the variable selection (Table 2, AUC=0.77). This model had better performance predicting Xer_{12m} than the reference model (Table 2), which was based on mean contralateral PG dose and $Xer_{baseline}$. No variable selection was performed for the reference model, thus internal validation was calculated by re-fitting the variables in the bootstrapped samples.

It should be noted that the variable Δ PG-volume was highly correlated to Δ PG-surface ($\rho=0.91$), and was therefore eliminated in the pre-selection. A model with Δ PG-volume and $Xer_{baseline}$ had comparable, but slightly reduced, performance measures (AUC=0.80; $R^2=0.34$; see supplementary material 3 for model and performance characteristics).

Parotid gland dose and surface change

A significant linear relation was observed between contralateral Δ PG-surface and contralateral PG dose ($p<0.001$), but the explained variance was relatively low ($R^2=0.27$). Based on the scatterplot (Figure 3A) a quadratic relation was fitted and it proved significantly better than a linear fit ($R^2=0.34$, $p<0.001$, ANOVA F-test).



This quadratic fit improved the fit even more for the ipsilateral PG (linear fit $R^2=0.19$; quadratic fit $R^2=0.38$, $p<0.001$, ANOVA F-test). Note that the mean dose levels received by the ipsilateral PG were larger (Figure 3B). The residuals of the regression models were reasonably normally distributed.

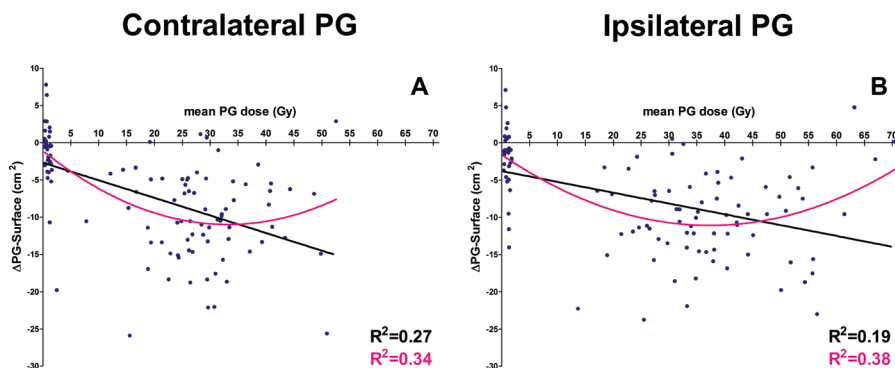


Figure 3 The relation dose and Δ PG-surface for both A) contra- and B) ipsilateral parotid gland (PG). Linear (black line) and quadratic (pink curve) regression curves were plotted.

Δ IBMs and other predictive variables

Initial (planning) PG mean dose did not significantly add to a model with Δ PG-surface in terms of predicting late Xer_{12m} (likelihood ratio test; $p=0.16$). Performance measures improved slightly, but no difference was seen after internal validation (Table 2).

A significant univariable logistic relation was found between Δ PG-surface and acute xerostomia scores at the same point in time, 6 weeks after radiotherapy ($Xer_{6w-post}$) ($p=0.017$; OR=0.93 (0.87 – 0.99); Figure 1-③). However, a stronger association between Δ PG Surface and Xer_{12m} was observed ($p<0.001$; OR=0.86 (0.79 – 0.93)) (Figure 1-①). Acute ($Xer_{6w-post}$) and late xerostomia (Xer_{12m}) were indeed related ($p<0.001$; OR=14.29 (5.20-39.27)) (Figure 1-④). Moreover, acute xerostomia added significantly to Δ PG-surface in predicting Xer_{12m} (likelihood ratio test; $p<0.001$) and vice versa (likelihood ratio test; $p<0.001$) (Figure 1-⑤). The performance measures of this model with Δ PG-surface and $Xer_{6w-post}$ further improved to an AUC of 0.90 (0.84 – 0.96) and R^2 of 0.56 (Table 2). Again, mean PG dose could not improve the model (likelihood ratio test; $p=0.27$). Calibration of all presented models was good (Table 2: Hosmer-Lemeshow test, calibration intercept and slope).

Table 2 (part 1). Model characteristics and performance measures of reference model, Δ IBM with and without dose models and combined acute xerostomia post-treatment model

		β		OR		p value
		Apparent	Corrected	Apparent	Corrected	Apparent
Reference model	Intercept	-2.613	-2.489*			
	Xer _{baseline}	1.570	1.500*	4.81	4.48*	0.004
	PG dose	0.049	0.047*	1.05	1.05*	0.001
Δ IBM models	Intercept	-2.180	-1.647			
	Δ PG-surface	-0.155	-0.126	0.86	0.88	<0.001
	Intercept	-2.875	-2.112			
	Xer _{baseline}	1.587	1.242	4.89	3.46	<0.001
	Δ PG-surface	-0.154	-0.121	0.86	0.89	0.002
Δ IBM and dose model	Intercept	-3.295	-2.389			
	Xer _{baseline}	1.533	1.175	4.63	3.24	0.002
	Δ PG-surface	-0.130	-0.100	0.88	0.91	0.004
Acute Xerostomia model	PG dose	0.028	0.022	1.03	1.02	0.161
	Intercept	-2.763	-1.982			
	Xer _{baseline}	1.540	1.164	4.66	3.20	0.004
Acute Xerostomia + Δ IBM model	Xer _{6w-post}	2.581	1.950	13.21	7.03	<0.001
	Intercept	-4.257	-3.111			
	Xer _{baseline}	1.469	1.110	4.34	3.03	0.012
	Xer _{6w-post}	2.598	1.963	13.44	7.23	<0.001
	Δ PG-surface	-0.169	-0.128	0.84	0.88	0.002

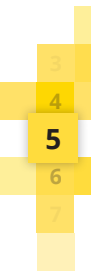


Table 2 (part 2). Model characteristics and performance measures of reference model, Δ IBM with and without dose models and combined acute xerostomia post-treatment model

	AUC		R ²		DS		HL X ² (p value)		Calibration intercept; slope	
	Apparent	Corrected	Apparent	Corrected	Apparent	Corrected	Apparent	Corrected	Apparent	Corrected
Reference model	0.76 (0.67-0.86)	0.75*	0.28	0.25*	0.21	0.20*	9.63(0.29)	0.20*	-0.01	-0.96
Δ IBM models	0.76 (0.66-0.85)	0.72	0.25	0.16	0.18	0.14	6.76(0.56)	0.14	-0.13	0.81
	0.82(0.72-0.91)	0.77	0.36	0.25	0.29	0.23	3.96(0.86)	0.23	-0.14	0.78
Δ IBM and dose model	0.82 (0.73-0.91)	0.77	0.38	0.26	0.30	0.24	8.08(0.43)	0.24	-0.14	0.77
Acute Xerostomia model	0.85 (0.77-0.93)	0.81	0.45	0.34	0.37	0.31	0.004 (1.00)	0.31	-0.11	0.76
Acute Xerostomia + Δ IBM model	0.90 (0.84-0.96)	0.86	0.56	0.44	0.45	0.39	5.12 (0.74)	0.39	-0.11	0.76

Abbreviations — PG: parotid gland; Xerbaseline and Xer6w-post: xerostomia score at baseline and 6 weeks after radiotherapy; Δ PG-surface: parotid gland surface difference after - before radiotherapy. Apparent: evaluated using the full dataset the same data. Corrected for optimism with bootstrapping the entire variable selection procedure; β : regression coefficients; OR: odds ratio; CI: confidence interval; AUC: area under the ROC curve; R²: Nagelkerke R²; DS: discrimination slope; HL: Hosmer-Lemeshow test. * No variable selection was performed for internal validation of the reference model.

Depicting the actual moderate-to-severe xerostomia incidences, figure 4 shows that Δ PG-surface was able to significantly differentiate between patients with high and low xerostomia incidence at 6 and 12 months.

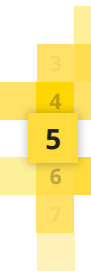
Using the complete early post-treatment model (Δ PG-surface, $Xer_{baseline}$ and $Xer_{6w-post}$) resulted in an even better distinction (Figure 4B), as the actual reported xerostomia differences of patients with high (>50%) and low (<50%) predicted risk were substantial.

Finally, using the same classification-based early post-treatment model for patients with moderate-to-severe xerostomia 6 weeks after radiotherapy ($Xer_{6w-post}=1$) showed that the predictions of the early post-treatment model could significantly differentiate between patients who recovered and those still suffering from xerostomia at 6 and 12 months (Figure 4C). This suggests that Δ PG-surface contributes in differentiating between patients that have persistent xerostomia up to 12 months and those that recover. Two patients had no reported xerostomia scores at 6 months.

Discussion

In this study, a significant relationship was shown between the geometric Δ IBM (Δ PG-surface) and late xerostomia. Δ PG-surface added significantly and independently to acute toxicities scores in predicting late xerostomia. Moreover, the performance of the models based on Δ PG-surface (with or without acute toxicities) were better than the reference model based on PG dose. Those observations together suggest that Δ PG-surface contains additional information on patient-specific development of late xerostomia. Mean PG dose did not add significantly to any of the Δ PG-surface models in this cohort. A possible explanation could be that Δ PG-surface and $Xer_{6w-post}$, which result from radiation dose, contain the same information as the PG mean dose, however this should be confirmed in an external dataset.

High correlation between Δ PG-volume and Δ PG-surface was observed. Prediction of late xerostomia was good with both variables, but Δ PG-surface performed better than Δ PG-volume (supplementary material 2). It can be hypothesized that surface change holds more information, because it also includes information on the shape of the PG. However, this observation may be limited to the current dataset, hence more research is necessary to investigate whether this can be confirmed in other datasets. Furthermore, in this study the absolute Δ PG-volume



and Δ PG-surface were investigated, similar performance was achieved with proportional change.

A non-linear (e.g. quadratic) relation between mean PG dose and Δ PG-surface (or volume) was observed, i.e. PG surface reduction increased with increasing the mean dose up to 30–40 Gy, but for PGs that received higher doses the PG surface reduction decreased again. This suggests that PGs react differently to higher doses, which might be due to direct necrosis of the PG cells inducing inflammatory swelling, instead of controlled apoptosis (21), which then compensates (partly) for the radiation-induced volume decrease.

Although there is no study to our knowledge that has investigated Δ PG-surface, many studies reported reductions of PG volumes after radiotherapy (22–27). In line with our results, the studies with adequate patient numbers, observed a significant, but weak, relation between mean PG dose and volume decrease ($r=0.41$ (24), $r=0.26$ (25)). This means that a large amount of unexplained variation remains. A possible explanation for these findings is variation in individual radiation sensitivity of PGs or mean PG dose may not be not the most optimal dosimetric parameter (9).

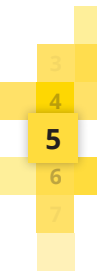
In the present study, significant associations were found between: Δ PG-surface and late xerostomia (Figure 1-①), the PG dose and Δ PG-surface (Figure 1-②) and PG dose and late xerostomia (Figure 1-‘reference model’). This is the first study that verified all these relationships simultaneously and developed a NTCP model to predict late xerostomia with a quantitative measure from CT imaging. Belli et al. showed a relation between PG shrinkage and acute xerostomia scores (28). Another study (25) observed in a limited cohort ($n \leq 24$, >12 months follow-up), that, small mid-treatment PG volume loss was associated with a longer time period of xerostomia recovery for patients receiving a relative high mean PG dose (>35.7 Gy, $n \leq 11$). These counterintuitive findings might be explained by the non-linear relation of the dose with Δ PG-surface shown in the current study. This relation suggests that a high PG dose may result in small Δ PG-surface (or volume), as this is potentially due to inflammatory PG swelling, which in turn might be related to the longer time period of xerostomia recovery. Hence, in this specific group of patients, Δ PG-surface alone might not optimally represent radiotherapy damage. In contrast, for patients with both a high PG dose and large PG change, Δ PG-surface still indicated high risk to develop late xerostomia in the current study.

PG surface reduction was related to acute xerostomia scores. However, a stronger relation was observed between PG surface reduction and late xerostomia (Xer_{12m}). Moreover, not only did acute xerostomia add predictive information to Δ PG-surface in predicting late xerostomia, also Δ PG-surface added to acute xerostomia. These results suggest that Δ PG-surface yields unique information about the patient-specific capability of the PG to recover from radiation damage (also see Figure 4C), and is not only a quantitative substitute of $Xer_{6w-post}$.

The model with Δ PG-surface, baseline and acute xerostomia scores (6 weeks after treatment) predicts late xerostomia with an exceptionally good performance, reflected in the good discrimination measures ($AUC=0.90$ (0.84–0.96) and $AUC_{bootstrapped}=0.86$). Early prediction of late xerostomia could improve effectiveness of future clinical studies, as the one year compliance is approximately 60% (1 year overall survival of HNC ~70% (29) together with other drop-out factors). An early surrogate could increase the follow-up information and thereby the time and cost effectiveness of clinical studies. Secondly, adequate prediction of late xerostomia can contribute to the physician–patient dialog, in order to discuss the chance of xerostomia recovery. Thirdly, selection of patients that do not recover from acute xerostomia (Figure 4C) can be beneficial for potential future treatments for xerostomia, such as stem cell therapy (9).

The relationship of Δ PG-surface with dose suggests that Δ PG-surface is a biomarker that measures physiological response. However, a correlation like this does not necessarily make this biomarker a surrogate for a clinical marker (30). A candidate surrogate marker should also have a relationship with the clinical endpoint, which is patient-rated late xerostomia in this study. This study shows a significant association between Δ PG-surface and Xer_{12m} . In addition, the model of Δ PG-surface together with acute xerostomia ($Xer_{6w-post}$) also meets the criteria mentioned above. Therefore, this model can be considered as a candidate surrogate marker for late xerostomia. Subsequently, external validation or a clinical trial is needed to verify whether the model of Δ PG-surface together with $Xer_{6w-post}$ can be used as a validated surrogate marker (30, 31).

Unfortunately, no contrast was used for the CT scan 6 weeks after treatment. Although this does not influence the geometric Δ IBMs, it could explain why no strong relation was observed between late xerostomia and CT-intensity based Δ IBMs, such as mean intensity/density change that has been reported in other studies (28). Univariable analysis, however, did show a significant relation



Moderate-to-severe xerostomia at different time points

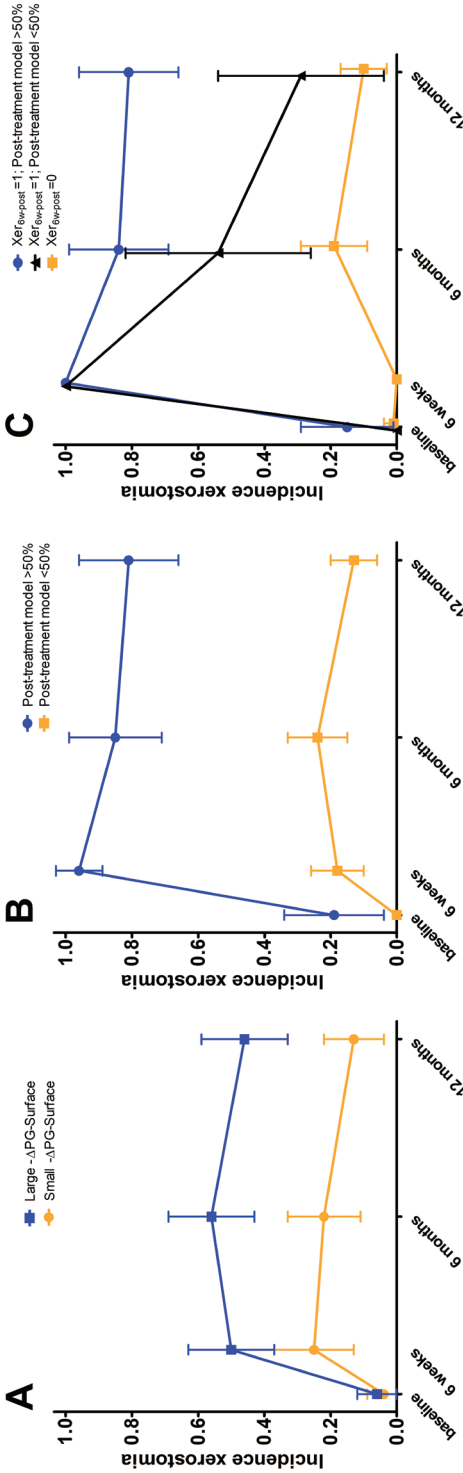
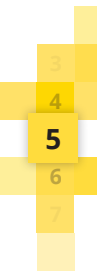


Figure 4 Actual moderate-to-severe xerostomia incidence and 95% confident intervals at baseline, 6 weeks, 6 months and 12 months after treatment for patients with: A) large and small PG surface reduction with threshold at the median (Δ PG-surface = -6.5 cm²), B) high (>50%) and low (<50%) risk predicted by early post-treatment model, C) high (>50%) and low (<50%) predicted risk for patient with moderate-to-severe xerostomia at 6 weeks after radiotherapy (Xer_{6w-post} = 1) and those without acute xerostomia (Xer_{6w-post} = 0).

between mean CT intensity and Xer_{12m} ($p=0.019$). Textural IBM changes (refer to supplementary material 4 for textural IBM details) were also tested in the current cohort, as described in a previous study (32). Univariable analysis showed that some textural IBMs were significantly associated with Xer_{12m} , however none gave a significant addition to ΔPG -surface. Textural IBM changes may yield similar information as ΔPG -surface or be biased due to the presence of metal artefacts in some patients. These IBMs were not extensively discussed in this study, since they gave no conclusive rejectable results ,due to the above discussed limitations. Furthermore, since the final models presented in our manuscript may be susceptible to limitations of the chosen variable selection procedure, LASSO regularisation, which is an alternative variable selection approach, was additionally performed and resulted in very comparable variable selection frequencies (ΔPG -surface was the most selected ΔIBM : 49% of the bootstrapped samples). This suggests a relatively large robustness of the associations found in this dataset, independent of the method of analysis. Additionally, modalities such as PET and MRI, that provide functional information could contribute in determining functionality loss of the PG gland, and could further improve quantifying and understanding the development of xerostomia.

Conclusion

Parotid gland surface reduction between start and 6 weeks after radiotherapy (ΔPG -surface) was significantly associated with the development of xerostomia 6 to 12 months after completing radiotherapy. Mean PG dose significantly correlated with ΔPG -surface , and did not add information to the ΔPG -surface model in predicting late xerostomia in this cohort. The model with ΔPG -surface and acute xerostomia early after radiotherapy significantly improved model performance to predict late xerostomia ($AUC=0.90$ ($0.84 - 0.96$); $AUC_{bootstrapped}=0.86$) and can therefore be a good candidate surrogate marker for late xerostomia at subsequent time points.



References

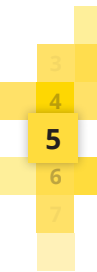
1. Langendijk JA, Doornaert P, Verdonck-de Leeuw IM, et al. Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. *J. Clin. Oncol.* 2008;26:3770–6.
2. Beetz I, Schilstra C, Van Der Schaaf A, et al. NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: The role of dosimetric and clinical factors. *Radiother. Oncol.* 2012;105:101–106.
3. Houweling AC, Philippens MEP, Dijkema T, et al. A comparison of dose-response models for the parotid gland in a large group of head-and-neck cancer patients. *Int. J. Radiat. Oncol. Biol. Phys.* 2010;76:1259–65.
4. Buettner F, Miah AB, Gulliford SL, et al. Novel approaches to improve the therapeutic index of head and neck radiotherapy: An analysis of data from the PARSPORT randomised phase III trial. *Radiother. Oncol.* 2012;103:82–87.
5. van Dijk LV, Brouwer CL, van der Schaaf A, et al. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. *Radiother. Oncol.* 2017;122.
6. Jensen K, Bonde Jensen A, Grau C. The relationship between observer-based toxicity scoring and patient assessed symptom severity after treatment for head and neck cancer. A correlative cross sectional study of the DAHANCA toxicity scoring system and the EORTC quality of life questionnaire. *Radiother. Oncol.* 2006;78:298–305.
7. Meirovitz A, Murdoch-Kinch CA, Schipper M, et al. Grading xerostomia by physicians or by patients after intensity-modulated radiotherapy of head-and-neck cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 2006;66:445–453.
8. Pringle S, Van Os R, Coppes RP. Concise review: Adult salivary gland stem cells and a potential therapy for xerostomia. *Stem Cells.* 2013;31:613–619.
9. van Luijk P, Pringle S, Deasy JO, et al. Sparing the region of the salivary gland containing stem cells preserves saliva production after radiotherapy for head and neck cancer. *Sci. Transl. Med.* 2015;7:305ra147.
10. Christianen MEMC, Schilstra C, Beetz I, et al. Predictive modelling for swallowing dysfunction after primary (chemo)radiation: results of a prospective observational study. *Radiother. Oncol.* 2012;105:107–14.
11. Christianen MEMC, Langendijk JA, Westerlaan HE, et al. Delineation of organs at risk involved in swallowing for radiotherapy treatment planning. *Radiother. Oncol.* 2011;101:394–402.
12. van der Laan HP, Christianen MEMC, Bijl HP, et al. The potential benefit of swallowing sparing intensity modulated radiotherapy to reduce swallowing dysfunction: an in silico planning comparative study. *Radiother. Oncol.* 2012;103:76–81.
13. Grégoire V, Levendag P, Ang KK, et al. CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC, RTOG consensus guidelines. *Radiother. Oncol.* 2003;69:227–236.

14. Vergeer MR, Doornaert PAH, Rietveld DHF, et al. Intensity-modulated radiotherapy reduces radiation-induced morbidity and improves health-related quality of life: results of a nonrandomized prospective study using a standardized follow-up program. *Int. J. Radiat. Oncol. Biol. Phys.* 2009;74:1–8.
15. Brouwer CL, Steenbakkers RJHM, Bourhis J, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother. Oncol.* 2015;117:83–90.
16. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat. Soc B.* 1995;57:289–300.
17. Van Der Schaaf A, Xu CJ, Van Luijk P, et al. Multivariate modeling of complications with data driven variable selection: Guarding against overfitting and effects of data set size. *Radiother. Oncol.* 2012;105:115–121.
18. Dehing-Oberije C, De Ruyscher D, Petit S, et al. Development, external validation and clinical usefulness of a practical prediction model for radiation-induced dysphagia in lung cancer patients. *Radiother. Oncol.* 2010;97:455–461.
19. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann. Intern. Med.* 2015;162:W1.
20. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: the R Foundation for Statistical Computing. 2011:Available online at <http://www.R-project.org/>.
21. Dewey WC, Ling CC, Meyn RE. Radiation-induced apoptosis: Relevance to radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* 1995;33:781–796.
22. Ajani A a, Qureshi MM, Kovalchuk N, et al. A quantitative assessment of volumetric and anatomic changes of the parotid gland during intensity-modulated radiotherapy for head and neck cancer using serial computed tomography. *Med. Dosim.* 2013:1–5.
23. Vásquez Osorio EM, Hoogeman MS, Al-Mamgani A, 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. *Int. J. Radiat. Oncol. Biol. Phys.* 2008;70:875–882.
24. Wang Z-H, Yan C, Zhang Z-Y, et al. Radiation-induced volume changes in parotid and submandibular glands in patients with head and neck cancer receiving postoperative radiotherapy: a longitudinal study. *Laryngoscope.* 2009;119:1966–1974.
25. Sanguineti G, Ricchetti F, Wu B, et al. Parotid gland shrinkage during IMRT predicts the time to Xerostomia resolution. *Radiat. Oncol.* 2015;10:15–20.
26. Broggi S, Fiorino C, Dell’Oca I, et al. A two-variable linear model of parotid shrinkage during IMRT for head and neck cancer. *Radiother. Oncol.* 2010;94:206–212.



Chapter 5 - Geometric ΔIBMs of the parotid gland are associated with late xerostomia

27. Reali A, Anglesio SM, Mortellaro G, et al. Volumetric and positional changes of planning target volumes and organs at risk using computed tomography imaging during intensity-modulated radiation therapy for head-neck cancer: an "old" adaptive radiation therapy approach. *Radiol. Med.* 2014;119:714–720.
28. Belli ML, Scalco E, Sanguineti G, et al. Early changes of parotid density and volume predict modifications at the end of therapy and intensity of acute xerostomia. *Strahlentherapie und Onkol.* 2014;190:1001–1007.
29. Pignon JP, Maître A le, Maillard E, et al. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): An update on 93 randomised trials and 17,346 patients. *Radiother. Oncol.* 2009;92:4–14.
30. Fleming TR, Powers JH. Biomarkers and Surrogate Endpoints In Clinical Trials. *Stat. Med.* 2013;31:2973–2984.
31. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *Eur. Urol.* 2015;67:1142–1151.
32. van Dijk LV, Brouwer CL, van der Schaaf A, et al. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. *Radiother. Oncol.* 2016:Article in press.



Chapter 6

Parotid gland surface area reduction during radiotherapy improves the prediction of late xerostomia

Under revision at: **Scientific Reports**

van Dijk LV, Langendijk JA, Zhai TT, Vedelaar TA, Noordzij W, Steenbakkers RJHM, Sijtsema NM.

Abstract

Background and purpose

Parotid gland response to radiation dose is patient dependent. The main aim of this study was to investigate whether parotid gland changes seen in weekly CT during treatment, quantified by delta image biomarkers (Δ IBMs), could improve the prediction of moderate-to-severe xerostomia at 12 months after radiotherapy (Xer_{12m}).

Materials and Methods

Patient-rated toxicity scores were prospectively collected. Parotid gland image characteristics (IBMs) at start and during treatment were extracted from planning and weekly CTs. The difference between these IBMs resulted in Δ IBMs, which represent geometric, intensity and texture changes of the parotid glands. Bootstrapped forward selection was performed to identify the best predictors of Xer_{12m} . The predictive contribution of the resulting Δ IBMs to a pre-treatment model, based on contralateral parotid gland mean dose (PGdose) and baseline xerostomia scores ($Xer_{baseline}$) only was evaluated.

Results

Moderate-to-severe xerostomia at 12 months (Xer_{12m}) was reported by 26 (38%) of the 68 patients included. The most predictive Δ IBM was the contralateral parotid gland surface change, which performed best for week 3 (Δ PG-surface_{w3}) and was significantly associated with Xer_{12m} ($p < 0.001$). Moreover, Δ PG-surface_{w3} showed a significant predictive contribution in addition to the pre-treatment model (likelihood-ratio test, $p = 0.002$), resulting a significantly better model performance (AUC=0.91) compared to that of the pre-treatment model (AUC=0.83).

Conclusion

The contralateral parotid gland surface area reduction between the 3rd week during and start of treatment (Δ Surface_{w3}) was associated with the development of late xerostomia. The mid-treatment Δ IBM model with Δ Surface_{w3} showed substantial predictive improvement over the pre-treatment model with PGdose and $Xer_{baseline}$ only.

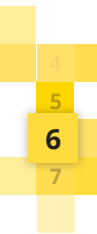
Introduction

Xerostomia is one of the most frequently reported side effects following radiotherapy of head and neck cancer (HNC) patients and affects patient-reported quality of life [1]. For the prediction of late xerostomia, Normal Tissue Complication Probability (NTCP) models have been developed with pre-treatment dose-volume parameters and baseline complaints as most important predictors [2,3]. However, xerostomia NTCP models based on information during treatment are less explored. Since in-treatment parameters contain information on the patient-specific response to treatment, they may resolve some of the unexplained variability that remains for NTCP models that are based on pre-treatment variables only. These in-treatment parameters could therefore be used to improve the prediction of late xerostomia. Adequate prediction supported by in-treatment data may offer new opportunities to guide treatment adaptation aiming at a further reduction of late radiation-induced side effects.

Several studies have investigated changes of the parotid glands during and after treatment in CT images [4–7] and have shown a weak to moderate relationship between parotid gland dose and volume change [4,6]. However, knowledge of the relationship between parotid gland changes and patient-reported xerostomia is still limited. Therefore, in our previous study, we investigated the association between late patient-reported xerostomia and parotid gland changes quantified in image biomarkers (IBMs) extracted from CT images before and 6 weeks after treatment. That study showed that the parotid gland surface reduction ($\Delta\text{PG-surface}_{6\text{w-postRT}}$) was strongly associated with the development of xerostomia at 6 and 12 months after radiotherapy [8].

However, this post-treatment model does not allow for treatment adaptation, as the total prescribed radiation dose has already been administered. Hence, the next step is to investigate parotid gland changes during treatment.

The aim of the current study was to identify quantitative parotid gland changes during treatment that predict the development of late xerostomia. These parotid gland changes were extracted from pre-treatment and weekly CT-images during radiotherapy, from which delta Image Biomarkers (ΔIBMs) were quantified, representing differences in intensity, texture and geometric characteristics of the parotid glands.



Method

Patients and image acquisition

The study cohort included consecutive HNC patients that were treated with definitive radiotherapy and received weekly CTs between January 2014 and December 2016. Radiation plans were adapted where necessary due to anatomical changes causing reduced target coverage. Patients were treated with IMRT or VMAT using a simultaneous integrated boost (SIB) technique, either as a single modality or in combination with concurrent chemotherapy or cetuximab. Plans were optimised to spare the parotid glands and swallowing organs at risk (superior pharyngeal constrictor muscle and supraglottic area) as much as possible without compromising the dose to the target volumes [9,10]. The primary tumour and pathologic lymph nodes were generally prescribed 70 Gy (2 Gy per fraction) and the cervical lymph node levels were prescribed an elective radiation dose of 54.25 Gy (1.55 Gy per fraction) [11]. More detailed descriptions of the radiation protocols used were reported in previous papers [12,13]. Patient characteristics are listed in Table 1.

Patients were excluded if they had salivary gland tumours, underwent prior surgery and/or underwent re-irradiation. An additional requirement was that patient-rated follow-up information at 6 and/or 12 months after radiotherapy was available.

CT scans (Somatom Sensation Open, Siemens, Forchheim, Germany; voxel size: 0.94 x 0.94 x 2.0 mm³; 100–140 kV) were acquired within 2 weeks prior to treatment (CT₀) and weekly during treatment (CT_{w1-6}), where CT_{w1} was generally acquired on the day of the first or second fraction. Patients only received intravenous contrast for CT₀. All scans were acquired with a thermoplastic mask in their radiotherapy treatment position. The data collection was part of routine clinical practice and therefore, the hospital ethics committee waived us from ethical approval.

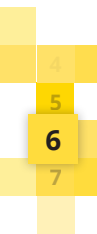
Endpoints

Patient-rated xerostomia scores were collected prospectively on a routine basis; before, weekly during, and subsequently 6 and 12 months after radiotherapy using the EORTC QLQ-H&N35 questionnaire, as part of the standard follow-up programme (SFP) (NCT02435576) [2,14]. The primary endpoint of this study was moderate-to-severe patient-rated xerostomia at 12 months after radiotherapy

Table 1 Patient characteristics of patients that had follow-up information available at 12 and 6 months after treatment.

Characteristics	Follow-up info at 12 months		Follow-up info at 6 months	
	N = 68	%	N = 88	%
Sex				
Female	20	29	26	30
Male	48	71	62	70
Age				
18-65	48	71	62	70
>65	20	29	26	30
Tumour site				
Oropharynx	22	32	27	31
Hypopharynx	0	0	1	1
Nasopharynx	5	7	5	6
Larynx	22	32	27	31
Oral cavity	15	22	23	26
Unknown primary	1	1	1	1
Other	3	4	4	5
Tumour classification				
T0	1	1	1	1
T1	11	16	14	16
T2	14	21	19	22
T3	17	25	19	22
T4	23	34	33	38
Unknown	2	3	2	2
Node classification				
N0	23	34	28	33
N1	9	13	13	15
N2abc	31	46	41	47
N3	3	4	4	5
Systemic treatment				
Yes	34	50	47	53
No	34	50	41	47
Treatment technique				
IMRT	27	40	30	34
VMAT	41	60	58	66
Bilateral				
Yes	57	84	72	82
no	11	16	16	18
Baseline xerostomia				
Any	26	38	36	41
None	42	62	52	59

Abbreviations: IMRT: Intensity-Modulated Radiation Therapy; VMAT: Volumetric Arc Therapy



(Xer_{12m}) and the secondary endpoint was moderate-to-severe patient-rated xerostomia at 6 months (Xer_{6m}). This corresponds to the 2 highest scores of the 4-point Likert scale (not, a bit, quite a bit, a lot).

Δ IBMs definitions

Parotid glands were delineated on the CT_0 according to the consensus guidelines of Brouwer et al. [15]. Delineations were warped to the weekly CTs using the deformable image registration tool in the treatment planning system RayStation v5.99 (RaySearch Laboratories, Stockholm, Sweden), the warped structures were manually corrected where necessary.

The image biomarkers (IBMs) were extracted from the planning and the weekly CTs with Matlab-based (Mathworks, Natick, MA, USA; version R2014a) in-house developed software. The definitions and formulas were in line with the 'Image biomarker standardisation initiative' [16]. The geometric IBM changes (geometric Δ IBMs) were calculated by subtracting the IBMs of CT_{w2-6} from those of CT_0 . This resulted in 15 geometric Δ IBMs per weekly CT, that for example represent volume, surface or compactness changes. Intensity IBMs describe first order and histogram characteristics of CT intensities of a parotid gland (e.g. mean or variance). Textural features describe the intensity heterogeneity and were extracted from the grey level co-occurrence matrix (GLCM) [17], grey level run-length matrix (GLRLM) [18,19], grey level size-zone matrix (GLSZM) [20] and neighbourhood grey tone difference matrix (NGTDM) [21]. Contrast enhancement was only used for CT_0 , and not for the weekly CT-scans. Since this can affect the intensity and texture IBM values, CT_{w1} (generally acquired before the 2nd radiation fraction) was considered to be the baseline CT. Hence, intensity and texture changes were quantified by calculating the difference between CT_{w1} and the CT_{w2-6} . As the intensity and texture IBMs can be influenced by metal artefacts, slices with metal artefacts were deleted and IBMs were calculated on the remaining slices only.

Figure 1 depicts the calculation of the Δ IBMs and CT time points. For a complete list of the 15 geometric, 17 intensity and 66 texture IBMs refer to the supplementary data 1, 2 and 3, respectively. Only Δ IBMs of the contralateral parotid gland were reported, as they performed better than those of the ipsilateral parotid gland. The geometric Δ IBMs were analysed separately from intensity and texture Δ IBMs.

Δ IBM selection

To identify the most predictive Δ IBMs, Δ IBM variable selection was performed each week. Firstly, Δ IBMs values were normalised by taking the difference between each value and the average, and then dividing by the standard deviation of the values of each Δ IBM. Secondly, a pre-selection that was based on inter-variable correlation was performed to reduce the number of variables. If the (Pearson) correlation of two variables was larger than 0.80, only the Δ IBM with the highest univariable association with the endpoint was selected [22]. Thirdly, stepwise forward selection was used to select the most important predictors (likelihood-ratio test: p-value <0.01) [23]. The entire variable selection procedure (normalisation, pre-selection and forward selection) was repeated on 1000 bootstrapped samples (i.e. with replacement), according to the TRIPOD guidelines [24]. The variable selection frequencies were evaluated to identify the most stable predictive variables per week. The Pearson correlation between the selected Δ IBMs was also investigated.

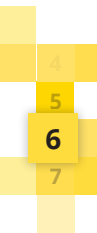
Δ IBM: univariable analysis

In order to identify the optimum week for predicting Xer_{12m} with Δ IBMs, the univariable associations were investigated for the selected Δ IBMs per week.

Δ IBM, dose and toxicity: multivariable analysis

A reference 'pre-treatment model' that was based on baseline xerostomia scores ($Xer_{baseline}$; none vs. any) and the contralateral PGdose, was fitted to the current dataset [2]. The prediction performance of the 'pre-treatment model' was first compared with that of models based on $Xer_{baseline}$ and the selected Δ IBMs. Subsequently, the addition of the selected Δ IBMs to the 'pre-treatment model' was investigated in terms of significance (likelihood-ratio test) and performance. Since our previous study showed that acute xerostomia scores significantly improved the Δ IBM model 6 weeks after treatment [8], we also investigated if the addition of acute toxicity as assessed during treatment to the Δ IBM-models improved model performance.

All multivariable models were logistic regression models. Model discrimination was measured with the area under the receiver operating characteristic curve (AUC) and the discrimination slope. Nagelkerke R^2 was used as a measure for explained variance. Model calibration was tested with the Hosmer–Lemeshow test and by repeating the entire variable selection on 1000 bootstrap samples, and



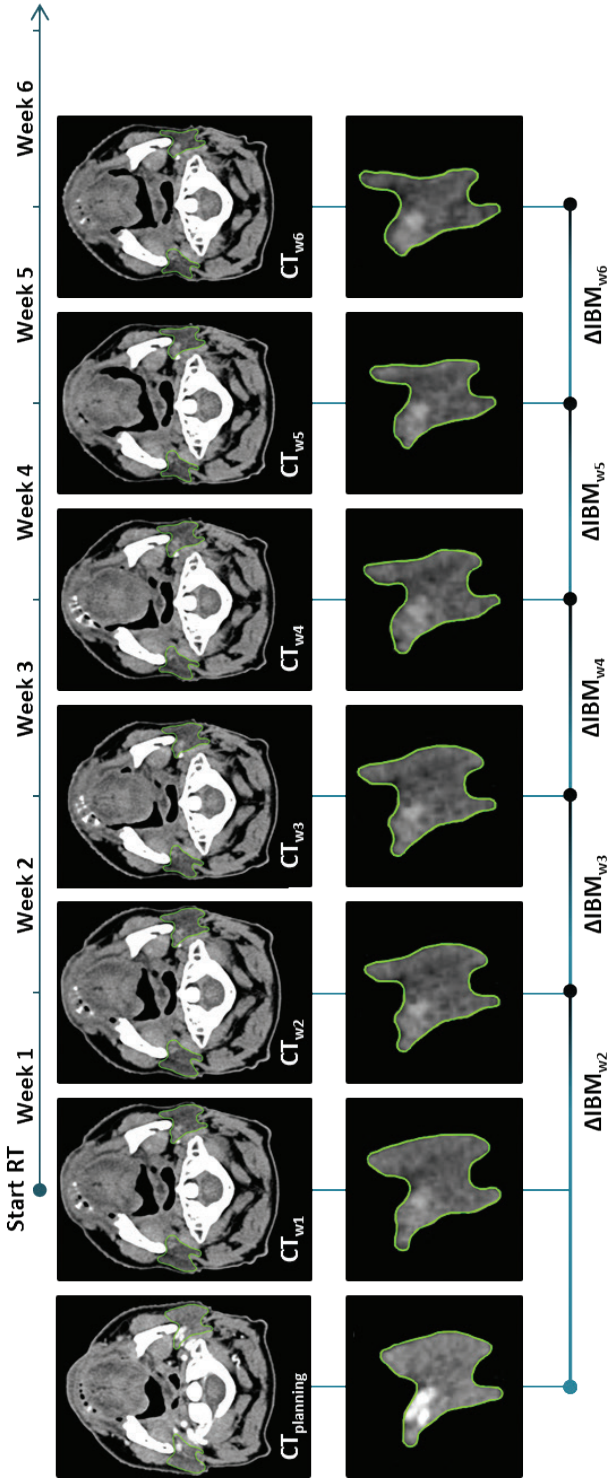


Figure 1 The difference between IBMs extracted from the weekly CTs (CT_{w_i}) and either the planning (geometric IBMs) or week 1 CT (intensity and texture IBMs) resulted in Δ IBMs per week. Generally, scans were taken at the start of consecutive radiation week.

by calculating the average of all resulting linear predictor slopes and intercepts. The coefficients were corrected for optimism according to this internal validation procedure.

The relationships between the resulting Δ IBM predictors, Xer_{baseline} , PGdose and acute xerostomia scores, were additionally explored with univariable logistic regression.

Parotid gland dose and Δ IBMs

The relationship of the mean contralateral PGdose and the Δ IBMs was investigated with linear regression. Model performance was measured with the coefficient of determination (R^2), and normality of the residuals of the regression models was checked.

Results

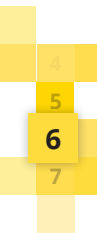
Patients

Moderate-to-severe xerostomia was reported by 26 (38%) of all 68 patients included at 12 months after radiotherapy. At 6 months after radiotherapy, the moderate-to-severe xerostomia reporting rate was 46 (52%) out of a total of 88 patients.

Δ IBMs selection

For the geometric image biomarkers, a change of contralateral parotid gland surface (Δ PG-surface) was the most frequently selected Δ IBM for all weeks predicting Xer_{12m} (see supplementary data 4 for frequency plot), except for week 6 where the Δ PG-volume frequency was slightly higher. The Δ PG-surface frequency was especially high in week 3 (882 selected in the 1000 bootstrap samples). In other weeks, the subsequently selected Δ IBMs, Δ 'volume', Δ 'bounding box volume' and Δ 'volume times mean intensity', were highly correlated with Δ PG-surface for all weeks with $\rho=0.77-0.94$, $\rho=0.63-0.82$ and $\rho=0.79-0.92$, respectively. Noteworthy, Δ PG-surface was also the most frequently selected Δ IBM for all weeks predicting the secondary endpoint Xer_{6m} .

For the intensity and texture IBMs, no clear selection of Δ IBMs that were most frequently selected for all weeks could be made (see supplementary data 4 for frequency plot). Overall, the most frequently selected Δ IBMs on average were: the large zone low grey level emphasis (LZLGE), coarseness from the neighbourhood



grey tone difference matrix (coarseness) and the median intensity (median). These Δ IBMs were moderately correlated to each other ($|\rho|=0.24-0.51$).

Δ IBM: univariable analysis

The univariable analysis also showed that Δ PG-surface was the most significant Δ IBM. This geometric Δ IBM was significantly associated with Xer_{12m} at all weeks ($p<0.04$) but was most significant in week 3 ($p<0.001$). This week showed the largest regression coefficient of all weeks (supplementary data 5).

For the intensity and texture IBMs, none of the most frequently selected Δ IBMs were significantly associated with Xer_{12m} in any of the weeks, yet the performance was best for week 3 Δ IBMs (LZLGE: $p\geq 0.06$; coarseness: $p\geq 0.06$; median: $p\geq 0.07$) (supplementary data 5).

Δ IBM, dose and toxicity: multivariable analysis

Since the Δ IBMs showed the best performance in week 3, the multivariable analysis was performed with the selected week 3 Δ IBMs only.

Discrimination of the reference 'pre-treatment' model ($Xer_{baseline}$ and PGdose) was good ($AUC=0.83$ ($AUC_{internal.val.}=0.82$)), yet the geometric Δ IBM model with Δ PG-surface_{w3} and $Xer_{baseline}$ performed better (Δ IBM model 1: $AUC=0.87$ ($AUC_{internal.val.}=0.86$)) in predicting Xer_{12m} (Table 2 and 3). Moreover, the addition of Δ PG-surface_{w3} to the pre-treatment model (likelihood-ratio test, $p=0.002$), significantly improved different aspects of performance (Δ IBM model 2: $AUC=0.91$ ($AUC_{internal.val.}=0.89$); Table 3).

Acute xerostomia scores at week 3 (Xer_{w3}) significantly improved this Δ IBM model 2 ($Xer_{baseline}$, PGdose and Δ PG-surface_{w3}) (likelihood-ratio test, $p=0.01$), but the improvement in performance was relatively small ($AUC=0.92$ ($AUC_{internal.val.}=0.90$)). The relationship between Δ PG-surface_{w3} and Xer_{w3} was not significant ($p=0.14$). This is also demonstrated in Figure 2, where patients with a large and small surface reduction at week 3 (Δ PG-surface_{w3}) showed a clear differentiation of actual moderate-to-severe xerostomia incidences at 6 or 12 months after treatment, but not for acute time points.

No significant relationship was found between Δ PG-surface and $Xer_{baseline}$ ($p=0.17$). Xer_{w3} was significantly associated with both PGdose ($p=0.04$) and $Xer_{baseline}$ ($p=0.03$), probably explaining the marginal prediction improvement of Xer_{3w} to the Δ IBM model with PGdose and $Xer_{baseline}$.

For the secondary endpoint Xer_{6m} , $\Delta PG\text{-surface}_{w3}$ also added significantly to the pre-treatment model in predicting Xer_{6m} (likelihood-ratio test, $p=0.02$). See supplementary data 6 for more details.

None of the frequently selected intensity or texture IBMs showed any significant improvement either compared to or in addition to the pre-treatment model (likelihood-ratio test, $p>0.27$) in predicting Xer_{12m} or Xer_{6m} .

Parotid gland dose and Δ IBMs

The linear relationship of contralateral parotid gland mean dose and $\Delta PG\text{-surface}$ was significant for all weeks ($p<0.008$; Table 4). Depicted in Figure 3, the regression coefficients of this linear relationship effectively increased over time, as did the coefficient of determination, but remained weak.

The selected intensity and the texture Δ IBMs, Δ median and Δ coarseness were significantly correlated ($p<0.05$) to parotid gland dose for week 2, 3, 5, 6 and 5, 6, respectively (supplementary data 7). However, the coefficient of determination was relatively low ($R^2=0.00\text{-}0.21$). $\Delta LZLGE$ was not significant for any week.

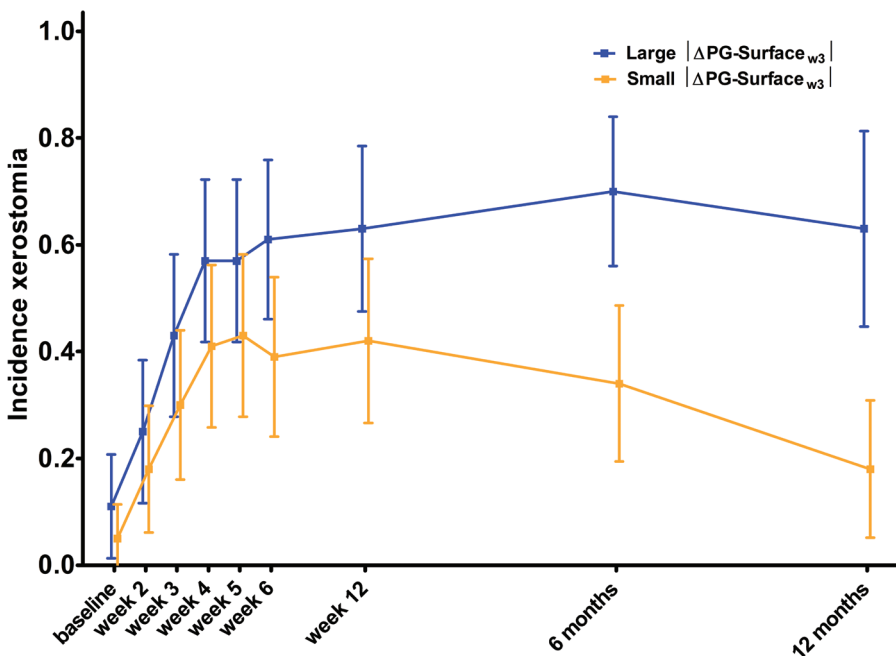
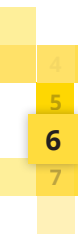


Figure 2 Actual moderate-to-severe xerostomia incidence and 95% confident intervals at baseline, weekly during, and 6 weeks (week 12), 6 months, and 12 months after treatment for patients, with parotid gland surface reduction at week 3 ($\Delta PG\text{-Surface}_{w3}$) larger (blue) or smaller (yellow) than the median reduction.



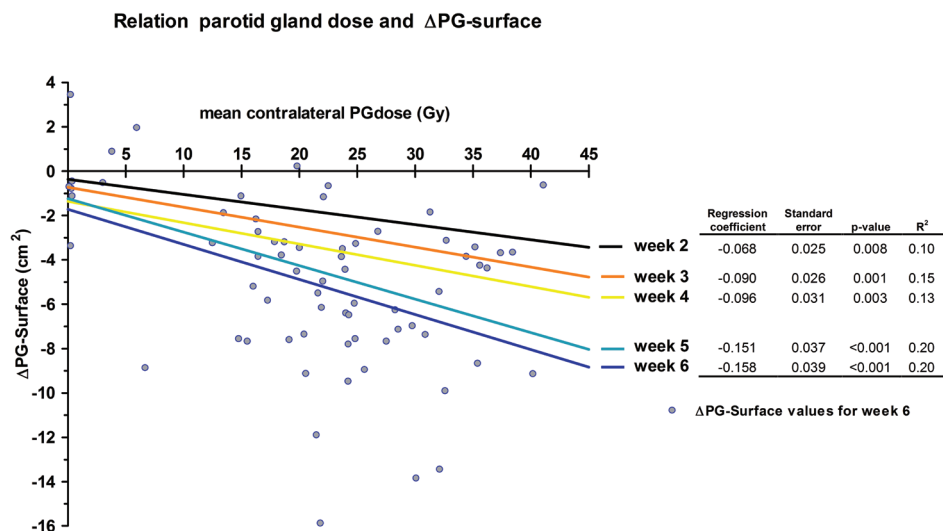


Figure 3 Univariable linear regression of contralateral parotid gland mean dose (PGdose) predicting parotid gland surface reduction (Δ PG-surface) for different weeks (lines) and regression characteristics (Table). Correlation increases over time, but remained weak. Data point represent Δ PG-surface values for week 6.

Table 2 Estimated coefficients (uncorrected and corrected for optimism) of pre-treatment and Δ IBM models

Model name		β		OR (95% CI)	p-value
		Uncorrected	Corrected		
Pre-treatment reference model	<i>intercept</i>	-3.794	-3.385*		
	Xer _{baseline}	2.531	2.280*	12.56 (3.39-46.54)	<0.001
	PG dose	0.099	0.089*	1.1 (1.03-1.18)	0.005
Δ IBM model 1	<i>intercept</i>	-3.139	-2.515		
	Xer _{baseline}	2.533	2.074	12.59 (3.13-50.73)	<0.001
	Δ PG-surface _{w3}	-0.568	-0.465	0.57 (0.41-0.79)	0.001
Δ IBM model 2	<i>intercept</i>	-4.515	-3.305		
	Xer _{baseline}	2.591	1.936	13.35 (3.13-56.95)	<0.001
	PG dose	0.072	0.054	1.07 (0.77-1.51)	0.074
	Δ PG-surface _{w3}	-0.481	-0.360	0.62 (0.57-0.67)	0.005

*No variable selection was performed for internal validation of the reference model
 Abbreviations: Xer_{baseline}: xerostomia at baseline ; PG dose: contralateral mean dose to parotid gland; Δ PG-Surface_{w3}: Parotid gland surface change from before and week 3 during treatment.
 β : regression coefficients; OR: odds ratio; CI: confidence interval; N.B. Surface change in cm²

Table 3 Performance of NTCP models IBMs

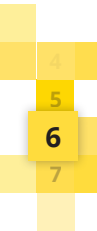
	Pre-treatment reference model	ΔIBM model 1	ΔIBM model 2
	Xerbaseline PG dose	Xerbaseline ΔPG-Surface w3	Xerbaseline PG dose ΔPG-Surface w3
Nagelkerke R ²	0.45	0.53	0.57
Area Under the Curve (AUC)	0.83 (0.72-0.94)	0.87 (0.79-0.96)	0.91 (0.85-0.98)
Discrimination slope	0.37	0.43	0.47
HL test X ² (p-value)	3.96 (p=0.41)	5.19 (p=0.27)	5.80 (p=0.21)
AUC _{corrected}	0.81*	0.84	0.87
Nagelkerke R ² _{corrected}	0.40*	0.45	0.46
Calibration slope (intercept)	0.90* (-0.03)	0.82 (-0.06)	0.75 (-0.07)

*No variable selection was performed for internal validation of the reference model
Abbreviations: HL: Hosmer–Lemeshow; corrected: corrected for optimism with bootstrapping;
IBM: Image Biomarker; Xerbaseline: xerostomia at baseline; PG dose: contralateral mean dose to parotid gland; ΔPG-Surface_{w3}: Parotid gland surface change from before and week 3 during treatment.

Discussion

The current study shows that surface change of the contralateral parotid gland (Δ PG-surface) assessed during the course of radiotherapy was strongly associated with the development of late xerostomia (Xer_{12m} and Xer_{6m}). The association of this geometric Δ IBM was statistically significant during the entire course of treatment but performed best for changes obtained between treatment planning and week 3. This time point is still clinically relevant, as any treatment adaptations could still influence the patient's toxicity outcome. Δ PG-surface_{w3} did not only show improved predictive performance over PGdose, but it also improved the pre-treatment model performance significantly. The resulting model that was based on $Xer_{baseline}$, PGdose and Δ PG-surface_{w3} showed excellent performance when predicting Xer_{12m} (AUC=0.91). However, these results should be confirmed by direct external validation.

Castelli et al. showed that parotid gland dose could significantly be reduced with an adaptive radiotherapy approach (ART) [25]. By re-planning the dose distribution on weekly CTs, an average NTCP reduction of 11% (maximum 30%) was observed. However, weekly re-planning is time consuming. This highlights



the potential of the Δ IBM NTCP model with Δ PG-surface_{w3}, since it could select patients during treatment that have a high risk of developing xerostomia. If these high-risk patients could, subsequently, receive less PGdose by re-planning, their risk of xerostomia could be further reduced. Alternatively, the model-based approach has been introduced to select patients for proton therapy. Patients can be selected that have a clinically relevant NTCP-reduction with a proton plan compared to their photon based treatment plan [26]. Proton therapy has the potential to better conform the dose to the tumour while sparing the surrounding normal tissue, due to the intrinsic properties of protons [27]. By incorporating patient-specific Δ IBM response information in the pre-treatment reference model, patients that do not initially qualify could be reclassified for proton therapy. Accordingly, treatment can be changed from photon to proton therapy, when relevant differences are seen in the new DNTCP values.

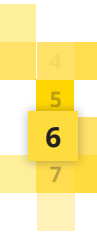
In a previous study, the geometric IBM differences were calculated between 6 weeks post-treatment and prior to treatment (Δ IBM_{6week-postRT}) [8]. The association of Δ IBM_{6week-postRT} with Xer_{12m} was investigated in a patient cohort (n=107) independent of the current cohort. Interestingly, the most stable and predictive post-treatment Δ IBM was also the contralateral Δ PG-surface. Similar to the results of the current study, inclusion of Δ PG-surface substantially improved the pre-treatment model. Using the same coefficients of the post-treatment model with Δ PG-surface_{6w-postRT}, Xer_{baseline} and PGdose in the current cohort, also showed a comparable performance (AUC=0.89) to that of the model trained in the current cohort (AUC=0.91). The other way around, using the coefficients of the current model in the previous cohort also resulted in a comparable improvement in performance for the model in the post-treatment cohort (supplementary data 8). This suggests that the Δ PG-surface_{w3} model would also perform well when externally validated in a cohort where Δ PG-surface is acquired at week 3. In both studies, Δ PG-volume was highly correlated to Δ PG-surface, and also performed well in predicting late xerostomia.

In line with other studies that observed a relationship between PGdose and PG shrinkage [4,6,28,29], linear regression in the current study also showed that there was a weak to moderate correlation between Δ PG-surface and PGdose. Interestingly, the correlation between PGdose and Δ PG-surface effectively increased over the time of treatment, illustrated by the increasing values of the regression coefficients and R² every consecutive week. This suggests that the effect of planning PGdose on Δ PG-surface becomes clearer as more dose is

administered. However, such an effect was not seen for the association between $\Delta\text{PG-surface}$ and Xer_{12m} , since the univariable logistic regression coefficient and the performance of $\Delta\text{PG-surface}$ increased from week 2 to 3, but decreased for the subsequent weeks. Hence, we concluded that the best moment for predicting Xer_{12m} was during week 3. The explanation may be that most parotid glands shrink when irradiated, as reported in previous studies [4–7], but patients that have a parotid gland that shrinks early in treatment have a higher risk of developing late xerostomia. Therefore, $\Delta\text{PG-surface}_{\text{week3}}$ could be a marker to differentiate between patients that develop permanent damage of the parotid gland versus those that can recover.

In addition to these observations, $\Delta\text{PG-surface}$ was not associated with acute xerostomia, although it was strongly associated with the development of late xerostomia. Figure 2 also demonstrated this, as $\Delta\text{PG-surface}_{w3}$ did not show a clear differentiation between the actual incidences of moderate-to-severe xerostomia at week 3 or any of the other acute time points. In contrast, this differentiation can be clearly seen for 6 and 12 months after radiotherapy. Furthermore, acute xerostomia scores at 3 weeks (Xer_{w3}) did significantly add to the model with $\text{Xer}_{\text{baseline}}$, $\Delta\text{PG-surface}$ and PGdose , although the improvement in performance measures was small. This is probably due to the correlation between Xer_{w3} and both PGdose and $\text{Xer}_{\text{baseline}}$. Further research needs to be performed on larger datasets in order to investigate whether acute toxicities can contribute to ΔIBM models.

Changes in intensity or texture IBMs were not related to the development of xerostomia. In contrast, many of these ΔIBMs were significantly related to PGdose , even though no relationship was seen with the development of xerostomia. Furthermore, detailed investigation of the most frequently selected intensity or texture ΔIBMs showed that these ΔIBMs contained one or two outliers that determined the effect. The influence of outliers indicates the importance of evaluating the selected IBMs before presenting them in a final model. In this study, the analysis of ΔIBMs was used rather than the IBMs directly extracted per week. The results of these absolute weekly IBMs were not significant. In contrast, in a previous study, a pre-treatment CT IBM that indicates tissue heterogeneity, was significantly associated with the development of late xerostomia [30]. It might be that the effect of pre-treatment is too weak to be observed in this relatively small dataset. In addition, using proportional ΔIBMs instead of absolute difference ΔIBMs did not improve the results of this study either.



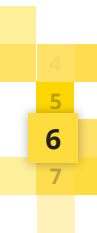
The limitations of this study are the low numbers of patients included in this analysis and no direct external validation was performed. Furthermore, only pre-treatment PGdose rather than the accumulated dose over all weekly CT scans was evaluated, since this was outside the scope of the paper. Brouwer et al. showed that accumulated dose calculated on weekly CTs was almost equal to the pre-treatment PGdose [31]. Using accumulated PGdose could improve the predictive performance of PGdose. Additionally, other modalities, such as positron emission tomography and magnetic resonance imaging could potentially provide better information during treatment on function loss of the PG gland. Future studies using these modalities could improve the quantification and understanding of the development of late xerostomia.

Conclusion

Contralateral parotid gland surface area reduction during the course of radiotherapy (Δ PG-surface) was associated with the development of late xerostomia both at 6 and 12 months after radiotherapy. The model consisting of $Xer_{baseline}$, parotid gland dose and Δ PG-surface, as assessed at week 3 during treatment (Δ PG-surface_{w3}), showed the best performance, and substantially improved the pre-treatment model based on parotid gland dose and $Xer_{baseline}$ only (from AUC of 0.83 to 0.91). This mid-treatment model may be a good candidate to identify patients most at risk of developing late xerostomia and who may benefit from treatment adaptations, but external validation is warranted.

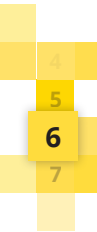
References

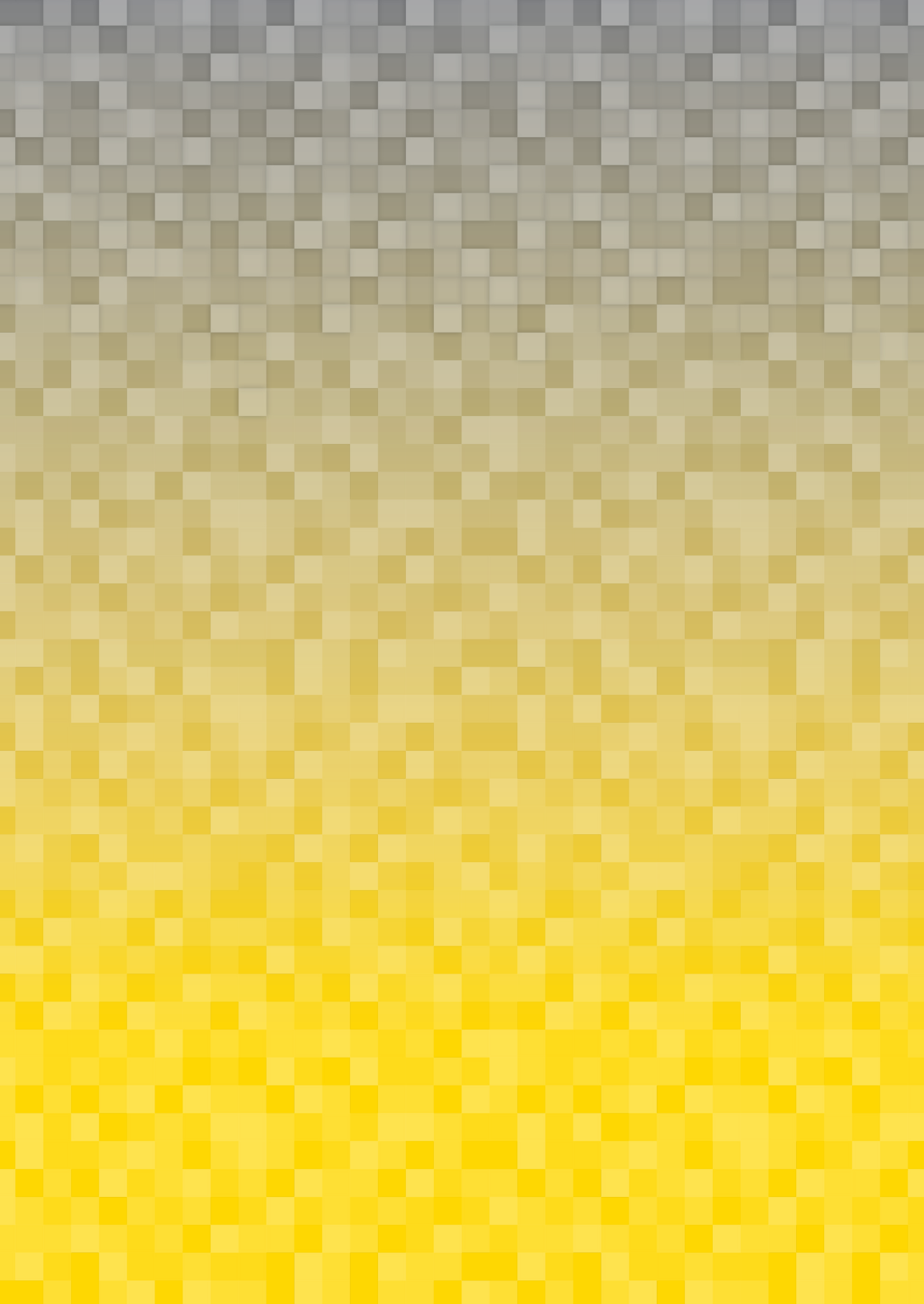
- 1 Langendijk JA, Doornaert P, Verdonck-de Leeuw IM, Leemans CR, Aaronson NK, Slotman BJ. Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. *J Clin Oncol* 2008;26:3770–6.
- 2 Beetz I, Schilstra C, Van Der Schaaf A, Van Den Heuvel ER, Doornaert P, Van Luijk P, 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. *Radiother Oncol* 2012;105:101–6.
- 3 Houweling AC, Philippens MEP, Dijkema T, Roesink JM, Terhaard CHJ, Schilstra C, et al. A comparison of dose-response models for the parotid gland in a large group of head-and-neck cancer patients. *Int J Radiat Oncol Biol Phys* 2010;76:1259–65.
- 4 Wang Z-H, Yan C, Zhang Z-Y, Zhang C-P, Hu H-S, Kirwan J, et al. Radiation-induced volume changes in parotid and submandibular glands in patients with head and neck cancer receiving postoperative radiotherapy: a longitudinal study. *Laryngoscope* 2009;119:1966–74.
- 5 Vásquez Osorio EM, Hoogeman MS, Al-Mamgani A, Teguh DN, Levendag PC, Heijmen BJM. Local Anatomic Changes in Parotid and Submandibular Glands During Radiotherapy for Oropharynx Cancer and Correlation With Dose, Studied in Detail With Nonrigid Registration. *Int J Radiat Oncol* 2008;70:875–82.
- 6 Reali A, Anglesio SM, Mortellaro G, Allis S, Bartoncini S, Ruo Redda MG, et al. Volumetric and positional changes of planning target volumes and organs at risk using computed tomography imaging during intensity-modulated radiation therapy for head–neck cancer: an “old” adaptive radiation therapy approach. *Radiol Med* 2014;119:714–20.
- 7 Brouwer CL, Steenbakkens RJHM, Langendijk JA, Sijtsema NM. Identifying patients who may benefit from adaptive radiotherapy: Does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help? *Radiother Oncol* 2015;115:285–94.
- 8 van Dijk L V., Brouwer CL, van der Laan HP, Burgerhof JGM, Langendijk JA, Steenbakkens RJHM, et al. Geometric Image Biomarker Changes of the Parotid Gland Are Associated With Late Xerostomia. *Int J Radiat Oncol Biol Phys* 2017.
- 9 Christianen MEMC, Langendijk JA, Westerlaan HE, Van De Water TA, Bijl HP. Delineation of organs at risk involved in swallowing for radiotherapy treatment planning. *Radiother Oncol* 2011;101:394–402.
- 10 van der Laan HP, Christianen MEMC, Bijl HP, Schilstra C, Langendijk J a. The potential benefit of swallowing sparing intensity modulated radiotherapy to reduce swallowing dysfunction: an in silico planning comparative study. *Radiother Oncol* 2012;103:76–81.
- 11 Grégoire V, Levendag P, Ang KK, Bernier J, Braaksma M, Budach V, et al. CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC, RTOG consensus guidelines. *Radiother Oncol* 2003;69:227–36.



- 12 Christianen MEMC, Schilstra C, Beetz I, Muijs CT, Chouvalova O, Burlage FR, et al. Predictive modelling for swallowing dysfunction after primary (chemo)radiation: results of a prospective observational study. *Radiother Oncol* 2012;105:107-14.
- 13 Van Der Laan HP, Gawryszuk A, Christianen MEMC, Steenbakkers RJHM, Korevaar EW, Chouvalova O, et al. Swallowing-sparing intensity-modulated radiotherapy for head and neck cancer patients: Treatment planning optimization and clinical introduction. *Radiother Oncol* 2013;107:282-7.
- 14 Vergeer MR, Doornaert PAH, Rietveld DHF, Leemans CR, Slotman BJ, Langendijk JA. Intensity-modulated radiotherapy reduces radiation-induced morbidity and improves health-related quality of life: results of a nonrandomized prospective study using a standardized follow-up program. *Int J Radiat Oncol Biol Phys* 2009;74:1-8.
- 15 Brouwer CL, Steenbakkers RJHM, Bourhis J, Budach W, Grau C, Grégoire V, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother Oncol* 2015;117:83-90.
- 16 Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative - feature definitions. *arXiv:161207003* 2016.
- 17 Haralick R, Shanmugan K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973;3:610-21.
- 18 Tang X. Texture information in run-length matrices. *IEEE Trans Image Process* 1998;7:1602-9.
- 19 Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process* 1975;4:172-9.
- 20 Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, et al. Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification. *Pattern Recognit Inf Process* 2009:140-5.
- 21 Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern* 1989;19:1264-73.
- 22 Van Der Schaaf A, Xu CJ, Van Luijk P, Van 't Veld AA, Langendijk JA, Schilstra C. Multivariate modeling of complications with data driven variable selection: Guarding against overfitting and effects of data set size. *Radiother Oncol* 2012;105:115-21.
- 23 Dehing-Oberije C, De Ruyscher D, Petit S, Van Meerbeeck J, Vandecasteele K, De Neve W, et al. Development, external validation and clinical usefulness of a practical prediction model for radiation-induced dysphagia in lung cancer patients. *Radiother Oncol* 2010;97:455-61.
- 24 Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* 2015;162:W1.
- 25 Castelli J, Simon A, Louvel G, Henry O, Chajon E, Nassef M, et al. Impact of head and neck cancer adaptive radiotherapy to spare the parotid glands and decrease the risk of xerostomia. *Radiat Oncol* 2015;10.

- 26 Langendijk JA, Lambin P, De Ruyscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. *Radiother Oncol* 2013;107:267–73.
- 27 Lomax A. Intensity modulation methods for proton radiotherapy. *Phys Med Biol* 1999;44:185–205.
- 28 Belli ML, Scalco E, Sanguineti G, Fiorino C, Broggi S, Dinapoli N, et al. Early changes of parotid density and volume predict modifications at the end of therapy and intensity of acute xerostomia. *Strahlentherapie Und Onkol* 2014;190:1001–7.
- 29 Broggi S, Fiorino C, Dell’Oca I, Dinapoli N, Paiusco M, Muraglia A, et al. A two-variable linear model of parotid shrinkage during IMRT for head and neck cancer. *Radiother Oncol* 2010;94:206–12.
- 30 van Dijk L V., Brouwer CL, van der Schaaf A, Burgerhof JGM, Beukinga RJ, Langendijk JA, et al. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. *Radiother Oncol* 2017;122:185–91.
- 31 Brouwer CL, Steenbakkens RJHM, van der Schaaf A, Sopacua CTC, van Dijk L V., Kierkels RGJ, et al. Selection of head and neck cancer patients for adaptive radiotherapy to decrease xerostomia. *Radiother Oncol* 2016:924–32.





Chapter 7

Summary and general
discussion

Summary and general discussion

This thesis is the first to show that image biomarkers (IBMs) can be used to improve prediction models for radiation-induced xerostomia in head and neck cancer (HNC) patients. These IBMs represent patient-specific tissue characteristics that are quantified in tangible values, allowing for quantitative analysis of three-dimensional clinical image information. We developed dedicated software to extract IBMs from clinical images. Toxicity prediction was improved by the addition of normal tissue IBMs (or Δ IBMs), which were either extracted before, during or after radiotherapy, to reference NTCP-models that were based on planning dose-volume parameters and baseline toxicity scores only. By optimizing toxicity prediction, this thesis contributes to the next step in personalized treatment approaches. Furthermore, it generated hypotheses for the patient-specific reaction to radiation dose, hereby advancing towards a better understanding of the development of late treatment-induced toxicities.

Part 1: Pre-treatment image biomarkers predict late toxicities

Chapter 2 reports the first published study on the association of salivary gland IBMs obtained from pre-treatment CT images with radiation-induced toxicities in HNC patients. The Short Run Emphasis (SRE) of the contralateral parotid gland was significantly associated to xerostomia 12 months after radiotherapy. This texture IBM added significantly to a model based on contralateral parotid gland dose and baseline xerostomia scores. Higher SRE values indicate a larger heterogeneity of the parotid tissue. Visual inspection of the CT images of patients with high and low SRE suggested that this heterogeneity was related to fat saturation of parotid glands, as infiltration of the fat tissue between the parenchymal gland tissue, resulting in increased texture in the gland due to the different image intensities of these tissues.

The resulting hypothesis was that the ratio of fat-to-functional parotid gland tissue may be an important pre-treatment factor in the development of xerostomia following radiotherapy.

In addition, maximum intensity within the submandibular glands seemed to be related to sticky saliva at 12 months after radiotherapy. This IBM was related to the CT intensity of the intravenous contrast of the artery or vein within the submandibular gland. However, next to the low stability of this IBM, no satisfying possible explanation could be found from the literature or the data for the

relation between this IBM and sticky saliva. More research is needed to interpret these findings.

Chapter 3 describes a study to investigate the added value of pre-treatment ^{18}F -FDG PET-IBMs in predicting late xerostomia. The PET-IBM that indicates the minimum value of the 90% highest SUVs (P90) was the most predictive of all intensity PET-IBMs. Additionally, the mean SUV also performed well, but P90 appeared more relevant in this dataset. Consequently, this study was the first to show that patients with low metabolic activity in the parotid glands were more likely to develop late xerostomia. The results of this study suggest that pre-treatment high metabolic activity in the parotid glands is associated with more viable parenchymal and/or stem cells with more repair capability and/or are less radiosensitive. Although, probably driven by other biological processes, the same is seen for high metabolic tumour tissue. Tumour tissue areas with high metabolic activity on ^{18}F -FDG PET images are where the recurrences are most likely to occur [1]. A possible explanation is that it arises from a combination of higher cell density, proliferation rate of metabolically active tissue and DNA repair capacity [2]. Considering that fat tissue is non-functional parotid gland tissue and is often low in intensity on ^{18}F -FDG PET images, the result of this study supports the hypothesis from the CT IBM study that the ratio of fat-to-functional parotid gland tissue may be an important pre-treatment factor for late radiation-induced xerostomia.

Building on the hypothesis that resulted from Chapter 2 and 3, **Chapter 4** describes parotid gland characteristics that were obtained with IBMs from Magnetic Resonance (MR) images. Since this modality is superior in imaging soft tissue contrast, it is more accurate in differentiating fat from the parenchymal gland tissue compared to CT and ^{18}F -FDG PET. MRI-based IBMs were significantly associated to xerostomia 12 months after treatment. Also in this study, the most robust and significant predictor was the P90, yet here indicating the 90th percentile of the standardized MR-intensities of the parotid glands. The MR-IBMs showed that high T1-signal within the parotid glands, which indicates high fat concentration, was significantly associated to higher rates of late xerostomia. Moreover, the addition of the MR-IBMs to the reference models significantly improved the prediction of xerostomia. The results of this MR-IBM study also supported the hypothesis that the ratio of fat-to-functional parotid gland tissue is an important predictor for late xerostomia.



Unrelated to the oncology field, previous studies have shown a relationship between fat saturation of the parotid glands and xerostomia related diseases: hyperlipidemia (elevated lipid plasma levels in the blood) and Sjögren's syndrome [3–5]. Izumi et al. [4] developed a MRI-based grading of the severity of parotid impairment for patients with Sjögren's syndrome that was based on similar image characteristics as where found by the studies of this thesis, i.e. high T1-weighted signal intensity areas (e.g. fat tissue) and heterogeneity in the parotid glands. Another study by Izumi et al. [5] also showed a relationship between increased signal intensities on T1-weighted MR images and impaired parotid function for patients suffering from hyperlipidemia. The findings of these studies suggest that increased fat concentration in the parotid gland can be caused by parenchymal changes due to lipid infiltration from the blood, and hereby reducing the functionality of the parotid gland. Since this is a continuous process, where only patients with severe hyperlipidemia experience xerostomia symptoms, pre-treatment oncology patients could have increased fat infiltrated parotid glands without having any symptoms (i.e. asymptomatic impaired parotid glands). These pre-treatment partially impaired parotid glands might have less reserve, less repair capacity and/or be more radiosensitive. This is in line with the results of Chapter 1, 2 and 3, which suggest that increased fat concentration in the parotid gland increases the probability of developing late xerostomia after radiotherapy. All these findings together lead to the hypothesis that high lipid blood levels are associated to the development of late radiation-induced xerostomia.

We recently initiated a study to investigate this relationship between blood lipid protein levels related to hyperlipidemia (e.g. triglycerides) and the development of late radiation-induced xerostomia. We aim to show a relationship between blood lipids and higher risk for patients to develop late xerostomia. This study will be the first to investigate the potential underlying pathophysiological process of toxicity development based on predictive image biomarkers. This will not only contribute to the improvement of personalised treatment approaches, but may also open doors for new xerostomia prevention research (e.g. prophylactic treatment of blood lipid reducers prior to treatment), due to a better understanding of the development of radiation induced xerostomia.

Another aspect that may be related to the increase of fat tissue in the parotid glands is ageing. With aging, the functional parotid tissue is substituted by connective tissue, such as fat or fibrous tissue in healthy individuals [1]. However, no age-related decrease of salivary flow rates is observed in the majority of

studies [1]. This suggests that the reserve capacity of healthy salivary glands is able to compensate for aging. After radiotherapy, typically no significant relationship between age and the development of radiation-induced xerostomia is found [2,3], which was also the case in the research of this thesis. This might be because age is an unreliable indicator of aging, since tissue aging rates fluctuate between individuals, as they are affected by multiple factors, such as lifestyle, environment and genetic factors [4]. Therefore, the effect of aging would be a more adequate variable, rather than age itself. The IBMs presented in this thesis may be a marker of the deterioration of functional parotid tissue as part of natural aging, which can be caused by lifestyle such as drinking or diet, or due to an underlying disease or a combination of both. As a consequence, the reserve capacity of the parotid glands is reduced, thereby the risk of radiation-induced xerostomia is increased.

IBM model applications (part 1)

One of the applications of prediction models is to guide treatment decisions. In radiotherapy, personalised treatment-decision making is gradually introduced. Very recently, the first step was made by using NTCP-models to decide whether patients should be treated with conventional photon radiotherapy or with proton therapy. Proton therapy (PT) was clinically introduced in January 2018 in the Netherlands, in the UMCG, providing an additional radiotherapy treatment option for HNC patients. This advanced treatment modality can limit radiation dose to spare normal tissues, while still delivering the prescribed dose to the tumour tissue [6–8]. Considering that the capacity for PT is limited, a model-based approach was introduced to select patients that are expected to benefit most from PT based on their toxicity risk reduction [9]. The model-based approach currently uses NTCP models that are based on dose-volume parameters and baseline toxicity scores. More accurate risk stratification could be obtained by using the pre-treatment IBM models presented in part one of this thesis (*Figure 1, application I*), but this requires external validation in independent datasets with sufficient numbers of patients.

The accurate pre-treatment NTCP estimation could also be used to select patients for other advanced treatment strategies (*Figure 1, application II*). Real-time MR-guided irradiation (MR-Linac) for example is introduced in radiotherapy, but, similarly to proton therapy, is limited available [10]. Since MR-Linac radiation



could reduce the error margins around the tumour, it may reduce surrounding normal tissue radiation.

Recent studies in the UMCG identified regions in the parotid glands containing high densities of stem cells and showed that sparing of those stem cells, which are presumed to be responsible for the parenchymal tissue regenerative ability, may reduce radiation-induced xerostomia [11]. Soon, the first clinical studies will start in which stem cells are extracted from patients before treatment and reintroduced after treatment to repair the regenerative ability of the salivary glands. The image biomarker xerostomia prediction models could guide the selection of patient that would benefit from such a treatment (*Figure 1, application III*). However, more research is necessary to determine if predicted high xerostomia risk patients would react well to future stem cell therapy.

Pre-treatment toxicity prediction could also be applied for model-based treatment plan optimisation (*Figure 1, application IV*). Instead of planning on dose-volume parameters, model-based optimisation uses NTCP values as objectives. NTCP values and dose-volume parameters cannot be directly translated, since their relationship is generally sigmoidal and not linear. Model-based optimisation is thus using weighted dose-volume parameters in order to plan directly on the patient-specific toxicities risk. IBM models could improve this approach even more, as IBMs quantify normal tissue characteristics, thereby taking into account the estimated tissue specific reaction to radiation dose. In other words, supplying the optimisation algorithm with more accurate toxicity estimation, allowing for more optimised organs at risk prioritisation.

Finally, for the current practise, where planning on dose-volume parameters is still the standard, treatment planners could also take into account that patients with unbeneficial parotid gland IBM values, such as high P90 in MR images or low P90 in ^{18}F -FDG PET images, are more likely to develop late xerostomia. For these specific patients, the sparing of salivary glands could be prioritized over other organs at risk, and hereby reducing the dose to the salivary glands.

Part 2: Image biomarkers changes after and during radiotherapy predict late toxicities

In the second part of this thesis, we investigated IBM changes (ΔIBMs) of the parotid gland after (Chapter 5) and during radiotherapy (Chapter 6) in relation to planning parotid gland dose and late xerostomia.

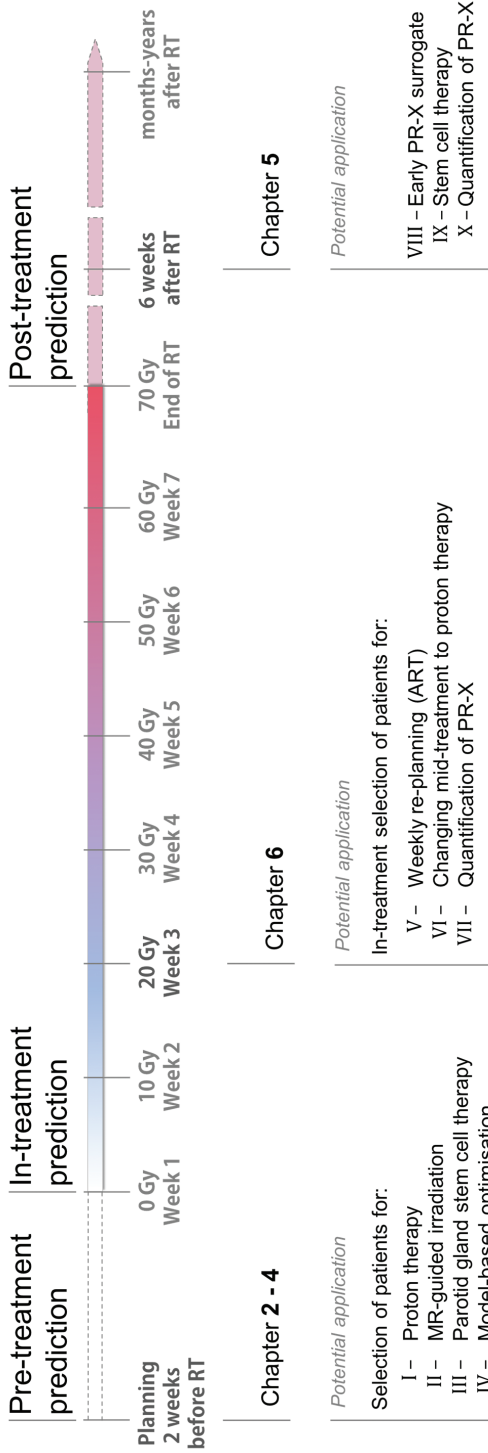


Figure 1 Schematic representation of IBM extraction time points: before (pre-treatment), during (in-treatment) and after (post-treatment) radiotherapy. The corresponded chapters and potential applications of models with IBMs to the different time points are listed.



In **Chapter 5**, Δ IBMs were extracted from post-treatment CT scans that were acquired 6 week after treatment and planning CT scans. Only geometric Δ IBMs were evaluated, since the post-treatment CTs were not contrast enhanced, unlike the planning CTs. The most predictive Δ IBM was Δ Surface of the contralateral parotid gland. This Δ IBM showed better and more robust performance than Δ Volume, yet these Δ IBMs were very strongly correlated to each other. It might be that Δ Surface partly quantifies shape deformation together with volume reduction. Furthermore, for patients that experienced moderate-to-severe xerostomia 6 weeks after radiotherapy, Δ Surface was also able to differentiate between patients that recovered and those that still experienced xerostomia at 6 and 12 months after treatment. In other words, this Δ IBM seems to provide a useful toxicity prediction early after treatment. The application of a post-treatment model are described in the 'IBM model applications' below. In addition, this study served as a proof of principle that Δ IBMs can predict the development of late xerostomia following radiotherapy.

Treatment options of radiation-induced xerostomia after radiotherapy are currently still limited. Therefore, the post-treatment model currently does not allow for interventions to reduce the risk of late xerostomia. Accordingly, we conducted a subsequent study using the data of an independent cohort of patients that received weekly CT scans as part of an adaptive radiation treatment scheme. In **Chapter 6**, the relationship is described between Δ IBMs, extracted from the weekly CTs, and the development of late xerostomia. Interestingly, the most predictive and frequently selected Δ IBM was again Δ Surface, which performed best at week 3 during treatment. The week 3 model with Δ Surface, parotid gland planning dose and baseline xerostomia scores had an excellent performance in selecting patients that are likely to develop late xerostomia. This time point offers opportunities for clinical interventions, as less than a third of the radiotherapy treatment has been completed.

IBM model applications (part 2)

The in-treatment IBM prediction, described in Chapter 6, can be used to guide adaptive radiotherapy (ART) approaches [12]. Since this model is developed for week 3 during radiotherapy, treatment adaptations can still be made that could influence the risk on xerostomia. In other words, early in-treatment prediction of normal tissue damage could potentially guide treatment adaptation, so that management of side effects can be improved or avoided. Castelli et al. showed

that the ART approach of weekly re-planning can reduce the parotid gland dose significantly [13]. They observed an average NTCP reduction of 11% (maximum 30%). However, since weekly re-planning is time consuming, early in-treatment selection of patients that would benefit from ART would be desirable (*Figure 1, application V*). Another application of the in-treatment IBM models could be to change treatment modality from photon to proton irradiation to limit the total parotid dose for patients that show a large Δ Surface at week 3 during treatment (*Figure 1, application VI*). A clinical trial will have to be conducted to evaluate if patients with a high risk of developing late xerostomia during treatment will benefit from the above-mentioned treatment adaptations.

The post-treatment model described in Chapter 5 could potentially serve as an early surrogate of late toxicity in the future (*Figure 1, application VIII*). Such an early surrogate of late xerostomia could improve effectiveness of future clinical studies, because the compliance of reporting side effects 6 weeks after treatment is much higher than a year after treatment (~ 60% 1-year compliance, due to death and other dropout factors). This could increase the cost-effectiveness of clinical studies. Early assessment of late xerostomia could also contribute to the physician-patient dialog, in order to discuss the probability of xerostomia recovery. Finally, the potential surrogate presented in this study, showed to discriminate between patients who do and do not recover from acute xerostomia. This may support the selection of high risk patients that may benefit from potential future xerostomia treatments, such as stem cell therapy [11], described in 'IBM model application (part 1)' (*Figure 1, application IX*). The post-treatment model can support the decision whether stem cell transplantation will be effective.

It is not surprising that Δ IBM models described in chapter 5 and 6 provided very adequate toxicity prediction, considering that meaningful Δ IBMs already capture a biological response of the parotid gland to radiation. Hence, Δ IBMs can also be regarded as a quantification of toxicities (*Figure 1, application VII and X*). The main aim of quantifying toxicity scores is to eliminate their subjective nature, since part of the unexplained variability of prediction models may be explained by this subjectivity. The individual experience of radiation-induced side effects with similar function loss varies widely among individual patients. This is illustrated by the discrepancy between physician-rated and patient-rated reporting of xerostomia scores [14]. However, the difficulty with validating the accuracy of



tissue damage quantification is that there is no reliable ground truth to compare it to, only the subjective measures.

Validation and future model development

A crucial next step for clinical implementation of the IBM NTCP models is external validation in larger patient cohorts. This would show that these models can also be utilised in new patient cohorts, and can be generally applied to obtain a more accurate prediction of late xerostomia than with conventional NTCP models. Therefore, the first next step should be to validate the pre-treatment models with new datasets from the UMCG and from other institutes. Datasets with multi-modality images (CT, PET and MRI) together with patient-reported xerostomia scores are sparse, but are becoming readily more available. In addition, the in-treatment model will need to be validated in a large cohort with late xerostomia scores and weekly CTs or high quality conebeam-CT scans. Furthermore, since the models' main aim is to select patients for more advanced treatment strategies, these models should in the future also be validated for the treatments they will be used for, such as proton therapy, MR-guided photon treatment and ART, as changes in treatment method can influence the NTCP models [15].

Implementation of prediction models is an iterative process, as models can be improved due to the availability of more knowledge and data over time. By evaluation of the predictive IBMs from all image modalities (CT, PET and MRI), the relationships between these IBMs should be explored more, and subsequently it can be evaluated whether the CT-, PET- and MR-IBMs add to each other in predicting late xerostomia or yield similar information, and the best IBM modality has to be determined. Based on the current studies, MR showed the most potential in predicting xerostomia, probably due to the good soft tissue contrast, yet we only investigated the simple sequences T1 weighted Turbo Spin Echo. Dedicated sequences that specifically differentiate fat from gland tissue, could improve the prediction further.

Ideally, treatment decision making is not only based on NTCP values, but also on tumour control probability (TCP) and/or survival probability. Multiple studies have observed a relationships between tumour image biomarkers and overall, disease-free and progression-free survival [16–21]. In work not included in this thesis, we showed that image biomarkers of the tumour tissue were able to predict overall survival better than clinical variables only [22]. Ultimately, treatment should ideally be optimised for every patient individually based on

tumour, normal tissue, patient and treatment characteristics, hereby, guiding patient-specific tumour and organ at risk dose objectives. For example, for a patient that has high estimated tumour control (e.g. based on HPV status and IBM tumour characteristics), but has also a high risk of developing late xerostomia, dose to the tumour might be somewhat compromised to decrease the dose to the parotid glands. The other way around, for a patient that has a high risk of recurrence and the NTCP values are low, intensification of the treatment (e.g. hyper-fractionation) might be considered. Nevertheless, more research has to be conducted to validate the models and test the feasibility of such approaches, before clinical trials can be designed to test dose limiting or intensification strategies based on image biomarkers.

Finally, the workflow that is presented in this thesis could be translated to other organ and tumour sites. Pre-treatment lung variations, quantified in image biomarkers, were shown to be related to radiation-induced pneumonitis in oesophageal cancer patients [23]. Similar to the work presented in this thesis the addition of predictive lung image biomarkers to conventional NTCP models could improve the prediction of toxicities.



References

- 1 Due AK, Vogelius IR, Aznar MC, Bentzen SM, Berthelsen AK, Korreman SS, et al. Recurrences after intensity modulated radiotherapy for head and neck squamous cell carcinoma more likely to originate from regions with high baseline [18F]-FDG uptake. *Radiother Oncol* 2014;111:360–5.
- 2 Roach MC, Turkington TG, Higgins K a, Hawk TC, Hoang JK, Brizel DM. FDG-PET assessment of the effect of head and neck radiotherapy on parotid gland glucose metabolism. *Int J Radiat Oncol Biol Phys* 2012;82:321–6.
- 3 Daskala ID, Tesseromatis CC. Morphological changes of parotid gland in experimental hyperlipidemia. *Int J Dent* 2011;2011.
- 4 Izumi M, Eguchi K, Ohki M, Uetani M, Hayashi K, Kita M, et al. MR imaging of the parotid gland in Sjögren's syndrome: A proposal for new diagnostic criteria. *Am J Roentgenol* 1996;166.
- 5 Izumi M, Hida A, Takagi Y, Kawabe Y, Eguchi K, Nakamura T. MR imaging of the salivary glands in sicca syndrome: comparison of lipid profiles and imaging in patients with hyperlipidemia and patients with Sjogren's syndrome. *AJR AmJRoentgenol* 2000;175:829–34.
- 6 Lomax A. Intensity modulation methods for proton radiotherapy. *Phys Med Biol* 1999;44:185–205.
- 7 van de Water TA, Bijl HP, Schilstra C, Pijls-Johannesma M, Langendijk J a. The potential benefit of radiotherapy with protons in head and neck cancer with respect to normal tissue sparing: a systematic review of literature. *Oncologist* 2011;16:366–77.
- 8 Van Dijk LV, Steenbakkens RJHM, Ten Haken B, Van Der Laan HP, Van't Veld AA, Langendijk JA, et al. Robust Intensity Modulated Proton Therapy (IMPT) increases estimated clinical benefit in head and neck cancer patients. *PLoS One* 2016;11.
- 9 Langendijk JA, Lambin P, De Ruyscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. *Radiother Oncol* 2013;107:267–73.
- 10 Lagendijk JJW, Raaymakers BW, van Vulpen M. The Magnetic Resonance Imaging–Linac System. *Semin Radiat Oncol* 2014;24:207–9.
- 11 van Luijk P, Pringle S, Deasy JO, Moiseenko V V, Faber H, Hovan A, et al. Sparing the region of the salivary gland containing stem cells preserves saliva production after radiotherapy for head and neck cancer. *Sci Transl Med* 2015;7:305ra147.
- 12 Castadot P, Lee JA, Geets X, Grégoire V. Adaptive Radiotherapy of Head and Neck Cancer. *Semin Radiat Oncol* 2010;20:84–93.
- 13 Castelli J, Simon A, Louvel G, Henry O, Chajon E, Nassef M, et al. Impact of head and neck cancer adaptive radiotherapy to spare the parotid glands and decrease the risk of xerostomia. *Radiat Oncol* 2015;10.

- 14 Jensen K, Bonde Jensen A, Grau C. The relationship between observer-based toxicity scoring and patient assessed symptom severity after treatment for head and neck cancer. A correlative cross sectional study of the DAHANCA toxicity scoring system and the EORTC quality of life questionnaire. *Radiother Oncol* 2006;78:298–305.
- 15 Beetz I, Schilstra C, Van Der Schaaf A, Van Den Heuvel ER, Doornaert P, Van Luijk P, 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. *Radiother Oncol* 2012;105:101–6.
- 16 Abgral R, Keromnes N, Robin P, Le Roux P-Y, Bourhis D, Palard X, et al. Prognostic value of volumetric parameters measured by (18)F-FDG PET/CT in patients with head and neck squamous cell carcinoma. *Eur J Nucl Med Mol Imaging* 2014;41:659–67.
- 17 Eary JF, O'Sullivan F, O'Sullivan J, Conrad EU. Spatial Heterogeneity in Sarcoma 18F-FDG Uptake as a Predictor of Patient Outcome. *J Nucl Med* 2008;49:1973–9.
- 18 Koyasu S, Nakamoto Y, Kikuchi M, Suzuki K, Hayashida K, Itoh K, et al. Prognostic value of pretreatment 18F-FDG PET/CT parameters including visual evaluation in patients with head and neck squamous cell carcinoma. *AJR Am J Roentgenol* 2014;202:851–8.
- 19 Alluri KC, Tahari AK, Wahl RL, Koch W, Chung CH, Subramaniam RM. Prognostic value of FDG PET metabolic tumor volume in human papillomavirus-positive stage III and IV oropharyngeal squamous cell carcinoma. *AJR Am J Roentgenol* 2014;203:897–903.
- 20 Zhai T-T, van Dijk L V, Huang B-T, Lin Z-X, Ribeiro CO, Brouwer CL, et al. Improving the prediction of overall survival for head and neck cancer patients using image biomarkers in combination with clinical parameters. *Radiother Oncol* 2017:256–62.
- 21 Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5.
- 22 Zhai T-T-T, van Dijk LVL V., Huang B-TB-T, Lin Z-XZ-X, Ribeiro COCO, Brouwer CLCL, et al. Improving the prediction of overall survival for head and neck cancer patients using image biomarkers in combination with clinical parameters. *Radiother Oncol* 2017.
- 23 Cunliffe A, Armato SG, Castillo R, Pham N, Guerrero T, Al-Hallaq HA. Lung texture in serial thoracic computed tomography scans: Correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development. *Int J Radiat Oncol Biol Phys* 2015;91:1048–56.



Nederlandse samenvatting

Introductie

Radiotherapie speelt een belangrijke rol bij de behandeling van hoofd-halskanker door de tumor lokaal te bestralen. In de afgelopen decennia is de levensverwachting van hoofd-halskanker patiënten na radiotherapie verbeterd. Hierdoor is er een groeiende groep overlevenden na kanker die langer leven met de bijwerkingen van de bestraling. Recentelijk zijn er nieuwe geavanceerde bestralingstechnieken zoals protontherapie ontwikkeld die schade in normaal weefselschade kunnen verminderen. Echter deze therapie is maar beperkt beschikbaar. Hierdoor wordt onderzoek naar het voorspellen van bijwerkingen als gevolg van bestraling door radiotherapie steeds belangrijker. Door beter te kunnen voorspellen en begrijpen welke patiënten een grote kans hebben op het ontwikkelen van bijwerkingen, kan de behandeling van hoofd-halskanker meer geoptimaliseerd worden voor de individuele patiënt. Een voorbeeld hiervan is om patiënten met het grootste risico op bijwerkingen te selecteren voor protontherapie.

De meest voorkomende bijwerkingen van hoofd-halsbestraling op lange termijn zijn droge mond (xerostomie) en taai slijm. Dit heeft een grote impact op de kwaliteit van leven. Om deze bijwerkingen te kunnen voorspellen zijn er predictiemodellen ontwikkeld die gebaseerd zijn op baseline klachten en de speekselklier bestralingsdosis. De performance van deze modellen is redelijk, maar om patiënten met een hoog en laag risico beter te kunnen onderscheiden moeten de modellen worden geoptimaliseerd. Ondanks dat patiënten dezelfde bestralingsdosis krijgen is er grote variatie in het ontwikkelen van late xerostomie en taai slijm.

De groeiende hoeveelheid beschikbare medische beeldvorming geeft, naast informatie over de tumor, ook veel ongebruikte extra informatie over de anatomie en fysiologie van de patiënt. We hebben software ontwikkeld om beeldkarakteristieken van weefsel te kwantificeren in zogenoemde image biomarkers (IBMs). IBMs geven informatie over intensiteit, textuur of geometrie van een bepaald interessegebied.

Het doel van dit onderzoek was om speekselklier disfunctionaliteit beter te voorspellen met behulp van IBMs. Deel 1 van dit proefschrift gaat over IBMs uit medische scans voor de radiotherapie en deel 2 uit beelden *na* en *tijdens* radiotherapie.

Deel 1: Image biomarkers (IBMs) voor radiotherapie voorspellen late xerostomie

Hoofdstuk 2 introduceert de eerste gepubliceerde studie waar IBMs worden gebruikt om bijwerkingen - als gevolg van gezonde weefselschade - te voorspellen. Het doel van dit onderzoek was om de predictie van xerostomie en taai slijm 12 maanden na radiotherapie te verbeteren met IBMs uit Computer Tomografie (CT) scans.

De CT-IBM die uit de variabele selectie voor xerostomie predictie werd geselecteerd, was de Short Run Emphasis (SRE) van de contralaterale parotisklier. De SRE was significant geassocieerd met late xerostomie en verbeterde het predictie '*referentie model*' dat gebaseerd is op de parotisklier bestralingsdosis en baseline xerostomie klachten. Hogere SRE-waarden zijn gerelateerd aan hogere heterogeniteit van het parotisklierweefsel. De visuele inspectie van parotisklieren met hoge en lage SRE op CT-scans suggereerde dat deze heterogeniteit gerelateerd was aan de infiltratie van vetweefsel tussen het functionele parenchym parotisklierweefsel. Dit resulteerde in de hypothese dat de verhouding van vet-tot-functioneel parotisklierweefsel een mogelijke risicofactor is voor de ontwikkeling van xerostomie na radiotherapie.

De maximale intensiteit van submandibularisklieren had een statistisch significante associatie met het ontwikkelen van taai slijm na radiotherapie. Deze IBM was gerelateerd aan de CT-intensiteit van het intraveneuze contrast in de submandibularisklier arteriën. Echter, de stabiliteit van deze CT-IBM is twijfelachtig. Meer onderzoek is nodig om deze bevindingen te kunnen interpreteren.

Hoofdstuk 3 beschrijft een onderzoek naar IBMs uit ^{18}F -FDG Positronemissietomografie (PET) scans voor het verbeteren van late xerostomie predictie. De best voorspellende PET-IBM was de P90 die de minimumwaarde aangeeft van de 10% hoogste SUV's. De gemiddelde SUV deed het ook goed, maar P90 leek relevanter in deze dataset. De resultaten van deze studie suggereren dat hoge metabolische activiteit in de parotisklieren gepaard gaat met meer functionele parenchymale cellen. Aangezien vetweefsel niet-functioneel parotisklierweefsel is en vaak een lage intensiteit heeft op ^{18}F -FDG PET-beelden, ondersteunen deze resultaten de hypothese dat de verhouding van vet-tot-functioneel parotisklierweefsel mogelijk een belangrijke risicofactor is voor late bestraling geïnduceerde xerostomie.



Hoofdstuk 4 bouwt voort op de hypothese van hoofdstuk 2 en 3, met parotisklier karakteristieken uit Magnetic Resonance Imaging (MRI) beelden. Doordat MRI een superieure modaliteit is op het gebied van weke delen contrast, kan MRI nauwkeuriger vet en parenchym klierweefsel differentiëren dan CT en ¹⁸F-FDG PET. Ook in deze studie was de P90 de meest robuuste en significante voorspeller, maar hier is het 90^{ste} percentiel van de gestandaardiseerde MR-intensiteiten van de parotisklieren. De MR-IBM toonden aan dat een hoog T1-signaal in de parotisklieren, wat wijst op een hoge vetconcentratie, significant geassocieerd was met hoger risico op late xerostomie. Bovendien, verbeterde de voorspelling van xerostomie aanzienlijk met de toevoeging van de MR-IBM aan de referentiemodellen. Ook deze resultaten ondersteunen de hypothese dat de verhouding van vet-tot-functioneel parotisklierweefsel een belangrijke voorspeller is voor late xerostomie.

Eerdere studies die niet aan hoofd-halskanker gerelateerd zijn, hebben een verband aangetoond tussen de vetverzadiging van de parotisklieren en xerostomia-gerelateerde ziekten: hyperlipidemie (verhoogde lipide plasma niveaus in het bloed) en het syndroom van Sjögren [3-5]. Soortgelijke beeldkenmerken als in dit proefschrift werden gepresenteerd als graderingsmarkers voor het syndroom van Sjögren: hoge T1-intensiteit en heterogeniteit van de parotisklieren [4]. Ook werd een verband aangetoond tussen verhoogde T1-intensiteit en verminderde parotisklierfunctie voor patiënten die lijden aan hyperlipidemie [5]. De bevindingen van deze studies suggereren dat verhoogde vetconcentratie in de parotisklier kan worden veroorzaakt door lipide-infiltratie vanuit het bloed, waardoor de functionaliteit van de parotisklier vermindert. Hoewel vetinfiltratie van de parotisklier kan gebeuren zonder symptomen, kan daardoor wel de reservecapaciteit worden verminderd waardoor de kans op late xerostomie na radiotherapie toch wordt verhoogd.

Samenvattend leidt dit tot de hypothese dat hoge niveaus van lipide in het bloed worden geassocieerd met een grotere kans op de ontwikkeling van late door straling geïnduceerde xerostomie.

Als vervolg op deze studies hebben we een onderzoek geïnitieerd om deze relatie tussen bloedlipide-eiwitniveaus gerelateerd aan hyperlipidemie (bijvoorbeeld triglyceriden) en de ontwikkeling van late door straling geïnduceerde xerostomie te onderzoeken. Dit is voor zover bij ons bekend, de eerste studie die op basis van IBMs het pathofysiologische proces van toxiciteitsontwikkeling onderzoekt.

Deel 2: Image biomarkers veranderingen na en tijdens radiotherapie voorspellen late bijwerkingen

In het tweede deel van dit proefschrift hebben we IBM-veranderingen (Δ IBM's) van de parotisklier na (hoofdstuk 5) en tijdens radiotherapie (hoofdstuk 6) onderzocht die gerelateerd zijn aan de ontwikkeling van late xerostomie.

In **hoofdstuk 5** werden Δ IBM's uit CT-scans voor en 6 weken na de behandeling verkregen. Het oppervlakteverschil van de contralaterale parotisklier (Δ Surface) was de meest voorspellende Δ IBM. Deze Δ IBM deed het beter dan Δ Volume, maar ze waren wel zeer sterk aan elkaar gecorreleerd. Mogelijk kan Δ Surface zowel de vervorming als de volumevermindering kwantificeren. Δ Surface was in staat onderscheid te maken tussen patiënten die wel of niet herstelden van ernstige xerostomie net na behandeling. Met andere woorden, Δ Surface kan snel na de behandeling een adequate lange termijn voorspelling geven. Dit kan nuttig zijn om een objectieve surrogaatmarker te hebben voor late xerostomie, en voor mogelijke toekomstige behandelingen van xerostomie.

De behandelingsopties van door bestraling geïnduceerde xerostomie zijn helaas op dit moment nog steeds beperkt. Daarom hebben we een volgende studie uitgevoerd in een onafhankelijk cohort van patiënten van wie wekelijkse CT-scans zijn verkregen als onderdeel van het adaptieve radiotherapie procedure waarin wekelijks bekeken wordt of het actuele bestralingsplan nog adequaat is..

In **Hoofdstuk 6** wordt de relatie beschreven tussen Δ IBMs, uit de wekelijkse CT's en de ontwikkeling van late xerostomie. Interessant is dat de meest voorspellende Δ IBM opnieuw Δ Surface was, die het beste voorspelde in week 3 tijdens de behandeling. Het model met Δ Surface van week 3, de parotisklier bestralingsdosis en de baseline xerostomie klachten presteerde uitstekend bij het selecteren van patiënten die mogelijk late xerostomie ontwikkelen. Dit tijdstip biedt kansen voor klinische interventies, aangezien minder dan een derde van de radiotherapiebehandeling is voltooid.



Dankwoord

Promoveren doe je voor jezelf, maar dat lukt niet zonder de mensen om je heen en dat is één van de hoofdredenen waarom ik onderzoek zo leuk vind. De expertise en enthousiasme van de mensen waarmee ik heb mogen werken, hebben mijn promotie traject extra leuk gemaakt. Daarnaast heb ik natuurlijk ook nog eens een prachtige groep mensen om me heen die me ook de nodige ontspanning geven naast het werk. Hieronder wil ik een paar mensen in het bijzonder bedanken.

Beste **Hans**, natuurlijk wil ik je bedanken voor al je ondersteuning en vertrouwen van de afgelopen jaren, waardoor niet alleen dit boekje tot stand is gekomen, maar ik me heb kunnen ontwikkelen als onderzoeker. Je denkt altijd mee met het onderzoek en je leest, ondanks je drukke schema, altijd mijn manuscripten met aandacht. Maar, je hebt het me niet altijd makkelijk gemaakt. Daarvoor moet ik je toch bedanken, want daardoor heb ik vaak mijn standpunten moeten overdenken, beter formuleren of aanpassen. Ik denk dat dit altijd geleid heeft tot beter onderzoek. Bedankt en ik kijk uit naar de research projecten die in de planning staan.

Beste **Roel**, tijdens mijn masterstage heb je me begeleid met mijn onderzoek, maar ook in de kliniek. Jouw manier van begeleiden heb ik altijd gewaardeerd, je gaf mij erdoor het zelfvertrouwen om de dingen die goed gingen, zelfstandig te doen. Door jou heb ik in korte tijd veel geleerd. Daarnaast, wil ik je ook bedanken dat je er voor me was als het af en toe minder ging en ik er even doorheen zat. Bedankt voor alle ervaringen die je met me hebt gedeeld, ik zal die altijd met me meedragen.

Beste **Marianna**, bedankt voor je geduld en je advies. Elke keer als ik weer een nieuwe versie van een manuscript had, keek je er altijd kritisch naar en gaf je me opbouwende opmerkingen. Je hebt me erg gesteund wanneer het schrijven of de resultaten tegen vielen en gaf me dan het vertrouwen dat het wel goed zou komen. Ook op persoonlijk vlak heb ik veel van je geleerd omdat we zo verschillen van karakter. Dank je wel. Daarnaast hebben we twee keer een super tijd gehad in bij de Radiomics meeting in Florida. Zee-vissen was echt super!

Dear **Tian Tian**, while sitting next to me for the majority of my PhD, you have not only been a supportive fellow PhD, but also a true friend with whom I feel that I can discuss anything. You always took the time, when I really needed it, to advise me, laugh with me, read whatever I wrote and listen to me. Thank you for all our conversations and experiences that we shared and I sincerely hope that I will also see you in China in a couple of years when you go back.

Beste **Lisa**, als Vlaamse heb je even moeten wennen aan mijn 'Nederlandse' directheid, maar ik ben blij dat je door de zure appel heen hebt gebeten en dat we zulke goede vriendinnen zijn geworden. Je weet me keer op keer te verrassen met jouw unieke enthousiasme, zoals die over het nationaal alarm of over die gekke podcast, waardoor ik weer anders naar de dingen ga kijken. Het maakt voor jou niet uit waarover, bij jou kan ik met al mijn werk, onderzoek en persoonlijke dilemma's terecht. Dank je wel!

Lieve **Veerle**, wat zijn wij fit geworden door al dat HIT en boxen, hè! Met een soeppie eten op maandag en op woensdag proberen je niet te schoppen tijdens het boxen, ben echt ik heel blij met zo'n lieve vriendin in Groningen. Want als jij iemand toelaat dan ben je er voor 100% als vriendin. Bedank dat je altijd bereid bent met me mee te denken.

Ik wil al mijn andere mede-PhDs, **Cassia**, **Lydia** en **Daan**, bedanken voor alle steun, gesprekken en het pingpongen in de pauze.

Beste **Wouter** en **Ruurd**, toen ik begon met mij promotietraject waren jullie mijn enige mede-PhDs; ik wil jullie voornamelijk bedanken voor het vele lachen.

De leden van de promotiecommissie wil ik bedanken voor het beoordelen van mijn werk en het plaatsnemen in de oppositie.

Dear **Antje**, hey coach! Thank for advising me in all different aspects of my research but also my career plans, by guiding and supporting me when I'm stuck on things. I'm very happy that you picked Groningen to work and that we can talk often.

Beste **Hans Paul**, vaak hebben wij kunnen sparren over mijn modelleer resultaten. Dit heeft me erg geholpen om te kunnen ontrafelen wat ze betekende en of ze wel logisch waren. Bij jou kon ik altijd terecht en hielp je mij door mijn standpunten te bevragen. Daarnaast hebben we ook vele goede gesprekken gehad als we even ons hoofd op andere zaken moesten zetten. Bedankt!

Lieve **Anniek**, vanaf het moment dat ik je als project manager leerde kennen op onze afdeling hadden we meteen een klik. Bedankt voor het meedenken en je enthousiasme bij het opzetten van nieuwe projecten, maar ook voor de feestjes en het sporten!

Lieve **Charlotte**, jij staat altijd positief tegenover nieuwe ideeën, en kan ze concretiseren. Bedankt voor al je steun tijdens mijn promotietraject, want voor al mijn vragen, zelfs over saaie administratieve dingen, nam je altijd de tijd.

Beste **Arjen**, bedankt voor het delen van je statistiek kennis, en je kon me altijd helpen als ik weer een moeilijke comment van de reviewers kreeg.

Beste **Peter** van Luijk, bedankt voor het speculeren over de mogelijk betekenis van de image biomarkers.

Lieve **AIOS** en **ANIOS**, waarvan er nu al een heel aantal radiotherapeut zijn. Ik wil jullie bedanken want ik heb me altijd welkom gevoeld bij jullie in de kliniek, de koffie en de vrimibo momenten. In het speciaal wil ik, **Crystal** bedanken voor je vriendschap en je humour, **Margriet** bedankt voor het introduceren van een hele andere wereld vol verbazing, we kunnen wel lachen hè, **Angelique** leuk dat je met Margriet en Crystal me kwamen opzoeken in New York, **Marloes** bedankt voor je geduld, liefde en de schoonheidsspecialiste-ervaring met de ampullen, **Anne** Niezink bedankt voor je enthousiasme, met jou samenwerken, vind ik extra leuk vanwege je andere kijk en aanpak.

Ik wil ook alle studenten, **Thea**, **Thije**, **Bill**, **Mohanned** en **Marije**, die ik heb of momenteel begeleid, bedanken voor de bijdrage aan projecten binnen en buiten dit proefschrift.

Beste **Jaap**, bedankt voor het klussen aan het MRI-kussen. Een aantal prototypen zijn voorbij gekomen, maar het is ons wel gelukt.

Lieve **Zwaanette** en **Miriam**, jullie hebben het merendeel van mijn PhD traject op mijn kantoor gezeten. Bedankt voor het gezellige kletsen, bucket-list details, steen-papier-schaar en de koffie.

Alle andere collega's, radiotherapeuten, klinisch fysici, laboranten, administratieve medewerkers, die ik niet bij naam heb genoemd maar wel mee heb samengewerkt, bedankt voor jullie hulp. Het is een fijne afdeling om in te werken als technisch geneeskundige en onderzoeker.

Van de Radiologie; **Jan Hendrik**, ik wil je bedanken voor al je geduld en tijd om een beetje van je kennis over MR sequenties en instellingen bij te brengen. Ik heb met veel plezier aan de DWI setting gesleuteld met jou. **Ronald Borra**, het is wat chaotisch samen werken met je, maar dat maakt het wel levendig, ik wil je bedanken voor je enthousiasme. **Hildebrand**, bedankt voor de DWI processing uitleg. De projecten lopen nog door, dus ik kijk uit naar de samenwerking in de toekomst.

Dear **Maria, Joe, Aditya, Rabia, Hyemin & Aditi**, thank you for your hospitality, the interesting projects and the great times I had with you in MSKCC and in New York.

Dear **Dave**, thank you for your enthusiasm in discussing potential research projects, sharing our data, mentoring me in my academic career and showing me and Tom Houston to the full extend. I'm very much looking forward to work with you in Houston.

Lieve vrienden, bedankt voor jullie humor en inspiratie: **Frank** voor je mooie verhalen, **Maaïke** voor je enthousiasme en doorzettingsvermogen, **Joyce** voor je lieve woorden, **Razmara** voor je gadgets wijsheid, **Solveig** voor je diepe gesprekken, **Maud** voor de leuke uitstapjes, **Sebas** voor het chillen, **Xanthe** voor de culturele uitstapjes, **de Bennies** voor mijn thuis buiten thuis, **Daphne** voor de spelletjes. Jullie zijn allemaal heel belangrijk voor mij.

Lieve **Loes**, mijn wijze zusje, die ondanks alles altijd lief voor me is en naar me luistert, want jij hebt het vermogen om echt te luisteren zonder oordeel. Cato Maté for life!

Lieve lieve **Kitty**, soms ken jij mij beter dan ik mezelf. In ieder geval weet je vaak beter wat goed voor me is. Dank je wel, dat je al 14 jaar mijn beste vriendin bent en direct vanaf het moment dat we elkaar leerden kennen mijn leven beter was omdat jij ervoor me bent, in alle omstandigheden. Voor mij ben je familie.

Lieve **Natasja** en **Merel**, mijn prachtig knappe paranimfen. Het is ongeloofelijk hoe ik veel ik met jullie heb kunnen en kan delen, dezelfde studie, het studentenleven, prachtige reizen, feesten, diepe gesprekken en ontzettend veel lachen. Allebei zijn jullie uniek op je eigen manier. Door jullie kan ik de wereld aan, ik kan niet zonder jullie.

Lieve **Renée** en **René**, wat zijn jullie lief voor me geweest tijdens mijn promotietraject, want soms kwam ik helemaal uitgeteld bij jullie aan en dan beurde Renée's heerlijke maaltijd me weer helemaal op, of dan kwamen jullie me ophalen van Schiphol. René bedankt voor het doorlezen van mijn Nederlandse teksten, toch een eer dat de editor van het AD mijn stukken heeft door gelezen (behalve dit dankwoord, ojee).

Lieve Eric en Brigitte, **papa** en **mama**, jullie geven je bodemloze en onvoorwaardelijk steun zo vanzelfsprekend dat ik nooit genoeg woorden heb om jullie daarvoor te bedanken, jullie zijn er gewoon altijd voor mij. Ik zou nooit zo mijn passie voor onderzoek hebben kunnen volgen en staan waar ik nu ben als jullie er niet zo voor me waren geweest. Ik ben trots op jullie, hoe jullie in het leven staan en hoe jullie me hebben opgevoed. Ik hou van jullie.

Mijn lieve **Tom**, wat ben jij er voor me geweest tijdens mijn PhD traject, en dat was soms niet makkelijk: laat thuiskomen, chaotische rommel, onbegrijpelijke enthousiasme, soms verdrietig, vaak weg en dan weer allemaal mensen over de vloer. Maar je laat me vrij en steunt me. Nu ben je al twee keer voor me verhuisd, eerst naar Utrecht en toen naar Groningen, maar straks gaan we samen, getrouwd en wel, het avontuur aan in Texas. Bedankt dat je mijn rots in de branding bent. Ik kijk uit naar de toekomst samen met jou.

Curriculum Vitae

Eighteen years after Lisanne V. van Dijk (Sanne) was born in 's Gravenhage (The Netherlands) on the snowy November 21th in 1988, she completed her VWO education and left her parental house in Zoetermeer to study Technical Medicine (TM) at the University of Twente (UT) in Enschede.

The TM bachelor and master 'Medical Imaging and Intervention' have started Sanne off with knowledge and expertise in both the medical and technical domain, training her to translate medical questions into technical solutions. A skillset in demand due to the increasing complexity of technical innovations in medicine. During Master internships, Sanne had the opportunity to observe and participate in different departments and hospitals in the Netherlands: Nuclear Medicine (UMCG, Groningen), Radiotherapy and Nuclear Medicine (NKI/AvL, Amsterdam), and Radiology (LUMC, Leiden). Her additional interest in particle physics, for which she has followed several additional related Applied Physics courses (UT), led her to the proton therapy facility (PSI) in Switzerland. Where her enthusiasm for proton therapy arose while researching proton beam specific margins. Subsequently, Sanne has continued researching robust proton therapy dose delivery for her master thesis project at the department of Radiotherapy at the UMCG.

With the PhD project presented in this thesis, she focused on personalized radiotherapy, predicting treatment outcome and side-effects with multi-modality imaging and complex modeling in head and neck cancer (HNC) patients. During this period, she also did research projects at MSKCC (New York), MAASTRO (Maastricht) and MDACC (Houston).

Since July 2018, Sanne is continuing her scientific ambitions as a post-doc at the Radiotherapy department (UMCG), where she will initiate projects to introduce image biomarker models into the clinic.



Grants and awards

- UMCG Kanker Research Fonds (KRF) (May, 2018), *High lipid blood levels as a risk factor for radiation-induced xerostomia*
- ESTRO Mobility Grant (June, 2018) *validation of image biomarker outcome prediction models* in MD Anderson
- DeepLearning challenge Award, QuantiVision IQwinterschool (March 7 – 9, 2018, Amsterdam)
- National clinical technical physician Award. Technical Innovations in Medicine (TiiM) Congres (2017)

Publications list

1. **van Dijk LV**, Steenbakkens RJHM, Ten Haken B, van der Laan HP, van 't Veld AA, Langendijk JA, Korevaar EW. *Robust Intensity Modulated Proton Therapy (IMPT) Increases Estimated Clinical Benefit in Head and Neck Cancer Patients*. **PLoS One** **2016**;11:1-14.
2. **van Dijk LV**, Brouwer CL, van der Schaaf A, Burgerhof JGM, Beukinga RJ, Langendijk JA, , Sijtsema NM, Steenbakkens RJHM. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. **Radiother Oncol** **2016**;122:185-191.
3. **van Dijk LV**, Thor M, Steenbakkens RJ, Apte A, Zhai T-T, Borra R, et al. Parotid gland fat related Magnetic Resonance Image biomarkers improve prediction of late radiation-induced xerostomia. **Radiother Oncol** **2018**. Accepted.
4. **van Dijk LV**, Brouwer CL, van der Laan HP, Burgerhof JGM, Langendijk JA, Steenbakkens RJHM, Sijtsema NM. Geometric image biomarker changes of the parotid gland are associated with late xerostomia. **Int J Radiat Oncol Biol Phys** **2017**:1101-10.
5. **van Dijk LV**, Noordzij W, Brouwer CL, Boellaard R, Burgerhof JGM, Langendijk JA, Sijtsema NM, Steenbakkens RJHM. ¹⁸F-FDG PET image biomarkers improve prediction of late radiation-induced xerostomia. **Radiother Oncol** **2017**;126:89-95.
6. **van Dijk LV**, Langendijk JA, Zhai TT, Vedelaar TA, Noordzij W, Steenbakkens, RJHM, Sijtsema NM. Parotid gland surface area reduction during radiotherapy improves the prediction of late xerostomia. **Scientific Reports** **2018**; Submitted
7. **van Dijk LV**, Langendijk JA, Sijtsema NM, Steenbakkens RJHM. Reply letter to "Texture analysis of parotid gland as a predictive factor of radiation induced xerostomia: A subset analysis". **Radiother Oncol** **2016**; 2017:322.
8. Schaake W, van der Schaaf A, **van Dijk LV**, Bongaerts AHH, van den Bergh ACM, Langendijk JA. Normal tissue complication probability (NTCP) models for late rectal bleeding, stool frequency and fecal incontinence after radiotherapy in prostate cancer patients. **Radiother Oncol** **2016**;119:381-7.
9. Brouwer CL, Steenbakkens RJHM, van der Schaaf A, Sopacua CTC, **van Dijk LV**, Kierkels RGJ, Bijl HP, Burgerhof JGM, Langendijk JA, Sijtsema NM. Selection of head and neck cancer patients for adaptive radiotherapy to decrease xerostomia. **Radiother Oncol** **2016**;0:924-32.

10. Beukinga J, Hulshoff J, **van Dijk LV**, Muijs CT, Burgerhof JGM, Kats-Ugurlu G, Slart RHJA, Slump CH, Mul VEM, Plukker JTM Predicting response to neoadjuvant chemoradiotherapy in esophageal cancer by textural features derived from pretreatment FDG-PET scans. **Ann Surg Oncol 2016**:S160.
11. Inoue T, Widder J, **van Dijk LV**, Takegawa H, Koizumi M, Takashina M, Usui K, Kurokawa C, Sugimoto S, Anneyuko I, Korevaar EW. Limited impact of setup and range uncertainties, breathing motion and interplay effects in robustly optimized intensity modulated proton therapy for stage III non-small cell lung cancer. **Int J Radiat Oncol Biol Phys 2016**:661–9.
12. Zhai T-T, **van Dijk LV**, Huang B-T, Lin Z-X, Ribeiro CO, Brouwer CL, Oosting SF, Halmos GB, Witjes MJH, Langendijk JA, Steenbakkers RJHM, Sijtsema NM. Improving the prediction of overall survival for head and neck cancer patients using image biomarkers in combination with clinical parameters. **Radiother Oncol 2017**:256–62.
13. van den Bogaard VAB, Ta BDP, van der Schaaf A, Bouma AB, Middag AMH, Bantema-Joppe EJ, **van Dijk LV**, van Dijk-Peters FBJ, Marteiijn LAW, de Bock GH, Burgerhof JGM, Gietema JA, Langendijk JA, Maduro JH, Crijns APG Validation and Modification of a Prediction Model for Acute Cardiac Events for Breast Cancer Patients Treated with Radiotherapy based on Three-dimensional Dose Distributions to Cardiac Substructures. **J Clin Oncol 2017**, 35(11), 1146–1148.
14. Huijsse SEM, Knopf A, **van Dijk LV**, Dierckx RAJO, Borra RJH. MR-only guided proton therapy : advances , future perspectives and challenges. **MReadings MR RT 2018**:2–7.
15. Apte AP, Iyer A, Crispin-ortuzar M, **Van Dijk LV**, Spezi E, Thor M, Um H, Veeraraghavan H, Oh JH, Shukla-Dave A, Deasy JO. Technical Note : Extension of CERR for computational radiomics: a comprehensive MATLAB platform for reproducible radiomics research. **Med Phys 2018**.
16. Schaake W, van der Schaaf A, van **Dijk LV**, van den Bergh ACM, Langendijk JA. Normal tissue complication probability (NTCP) models for late urinary incontinence, hematuria, pain and voiding frequency. **PLoS One 2018**. 13, 7, p. e0197757
17. Zhai TT, Langendijk JA, **van Dijk LV**, Halmos GB, Witjes MJH, Oosting SF, Noordzij W, Sijtsema NM, Steenbakkers RJHM. Prognostic Value of CT-based Image-biomarkers for Treatment Outcomes of Head and Neck Cancer Patients. In preparation

For a complete list: [Google Scholar](#) or [Research_gate](#)

Accepted abstracts:

1. Oral presentation. European Society for Radiotherapy & Oncology (ESTRO) conference (April 2018), Barcelona. *Fat-related image biomarkers improve radiation-induced xerostomia prediction*
2. Oral presentation. ESTRO conference (May 2017), Vienna. *¹⁸F-FDG PET image biomarkers improve prediction of late radiation-induced xerostomia*
3. Oral presentation. Conference Supporting health by Technology VII (May 2016): *Image characteristics improve the prediction of radiation-induced side-effects*
4. Oral presentation. ESTRO conference (April 2016), Turin: *CT Image biomarkers improve the prediction of xerostomia and sticky saliva*
5. Poster presentation. ESTRO conference (April 2016), Turin: *Early CT image biomarkers change and xerostomia score are strong predictors for late xerostomia*
6. Oral presentation. Annual PhD meeting Cancer Research Center Groningen (CRCG) (Oct 2015): *Early CT image biomarker changes during radiotherapy predict late xerostomia*
7. Oral presentation. ESTRO conference (April 2015): *Does the clinical benefit of IMPT persist if plans are made robust against setup and range errors?*