Next Generation Health Technology Assessment

# Project Title: Next Generation Health Technology Assessment to support patient-centred, societally oriented, real-time decision-making on access and reimbursement for health technologies throughout Europe

## Grant Agreement Number: 825162

# Deliverable Title: Reviews on applicable patients and models for proton therapy for patients with head and neck cancer

| | |
|---|---|
| Deliverable Nº 1.6 | Version: 3.0 |
| Date: 15/03/2022 | Lead Beneficiary: UMCU, UMCG, EORTC |
| Nature: Report | Diss Level: Confidential |

# TABLE OF CONTENTS

# DOCUMENT INFORMATION

| | | | |
|---|---|---|---|
| Grant Agreement Number | 825162 | Acronym | HTx |
| Full title | Next Generation Health Technology Assessment to support patient-centred, societally oriented, real-time decision-making on access and reimbursement for health technologies throughout Europe | | |
| Project URL | www.htx-h2020.eu | | |
| EU Project officer | LIMA DA CUNHA Carlos Eduardo Carlos-Eduardo.LIMA-DA-CUNHA@ec.europa.eu | | |

| | | | | |
|---|---|---|---|---|
| Deliverable | Number | 1.6 | Title | Reviews on applicable patients and models for proton therapy for patients with head and neck cancer |
| Work package | Number | 1 | Title | Treatment pathways in specific therapeutic areas |

| | | | | |
|---|---|---|---|---|
| Delivery date | Contractual | 31/12/2020 | Actual | 15/03/2022 |
| Status | | Version: | Draft ☐ Final ☑ | |
| Nature | | Report ☑ Prototype ☐ Other ☐ | | |
| Dissemination Level | | Public ☑ Confidential ☐ | | |

| | | | | |
|---|---|---|---|---|
| Authors (Partner) | Hans Langendijk (University Medical Center Groningen)<br>Ewoud Schuit (University Medical Center Utrecht) | | | |
| Other contributors | Lifang Liu (European Organisation for Research and Treatment of Cancer)<br>Junfeng Wang (Utrecht University) | | | |
| Responsible partner | Email | j.a.langendijk@umcg.nl; e.schuit@umcutrecht.nl | | |
| | Partner | UMCG<br>UMCU | Phone | +31 631 11 80 82<br>+31 503 61 55 32 |

| | |
|---|---|
| Description of the deliverable | The current report describes two literature reviews:<br>Task 1.4.1: A review on the effectiveness of proton therapy versus photon therapy for patients with head and neck cancer<br>Task 1.4.2: A review on prediction models that predict radiation-induced toxicities in patients with head and neck cancer |
| Key words | Head and neck cancer; HNC; literature review; systematic review; proton therapy; photon therapy; radiotherapy; normal tissue complication probability; NTCP; prediction model |

# DOCUMENT HISTORY

| NAME | DATE | VERSION | DESCRIPTION |
|---|---|---|---|
| Ewoud Schuit | 14/12/2021 | 1.0 | First draft |
| Hans Langendijk | 19/12/2021 | 2.0 | Second draft |
| Review SC members | 22/12/2021 | 2.0 | Final draft |
| Lifang Liu and Junfeng Wang | 15/03/2022 | 3.0 | Final |

## RATIONALE OF THE DELIVERABLE

Radiotherapy plays a pivotal role on locoregional control and overall survival of patients with head and neck cancer, whereas it may result in a broad spectrum of acute and late radiation-induced side effects, which negatively affect quality of life of patients.

Pencil beam scanning proton therapy (PBS-IMPT) is the alternative treatment modality to the conventional photon therapy applied to patients with head and neck cancer to reduce side effects without compromising treatment efficacy, such as local/regional control and survival.

Thus far, the reimbursement decision of proton therapy in the Netherlands is based on multivariable prediction models which predict a patient's normal tissue complication probability (NTCP) based on patient, disease, and treatment characteristics, including radiation dose to healthy tissues.

In alignment of the scope of the HTx project, it is crucial to understand the state-of-art on the efficacy of proton therapy and to identify the most reliable NTCP models based on assessment of quality and applicability. Therefore, the Case Study 1 starts with two literature reviews as described below.

# INTRODUCTION

Next Generation Health Technology Assessment (HTx) is exploring how to enhance methods for integrating evidence from randomized clinical trials (RCTs) and real-world data (RWD). Based on four relevant case studies, HTx is expanding statistical and econometric methods for generating robust estimates of effectiveness and cost-effectiveness in order to support relevant HTA decision-making for these complex and personalised combinations of health technologies. In addition, HTx will contribute to improving methods to support personalised treatment advice fitted for sharing with patients and their physicians. One of the four case studies is the model-based approach for proton radiation therapy (PRT) for head and neck cancer (HNC).

Since the characteristics of patients with cancer that arises in the head or neck region (in the nasal cavity, sinuses, lips, mouth, salivary glands, throat, or larynx) and tumour characteristics differ a lot, it is interesting to compare patients groups to see which patients will be benefit the most getting proton therapy. HNC is one of the cancers for which proton may be most beneficial, since in HNC patients target volumes are surrounded by numerous organs at risk (OARs). (Part 1 of this report: A review on the effectiveness of proton therapy versus photon therapy for patients with head and neck cancer) Proton radiation deposits most of its initial energy at the end of the range (Bragg peak) within the tumour. The use of more precise proton radiation techniques may reduce the dose in OARs as much as

possible and thereby reduce radiation-induced side effects that typically strike HNC patients. In patients with HNC, OARs include i.e. the salivary, parotid, submandibular, sublingual glands, and thyroid glands. Radiation of these OARs can result in subsequent complications and may have a significant impact on health-related quality of life (HR-QoL).

The OARs are evaluated mathematically in so-called multivariable normal tissue complication probability (NTCP) models, which also consider the most relevant dose-volume parameters as well as other independent prognostic factors, such as age. The outcome of a NTCP model is the chance that healthy tissue will be damaged by the radiation treatment. Estimation of the dose-volume-effect or NTCP of critical organs is an essential factor prior to the delivery of radiotherapy, because very often critical organs, within the vicinity of the tumour, receive a radiation dose equal or less to that of the tumour. The NTCP model is adjusted per complication, resulting in the five different NTCP models for the most relevant complications sticky saliva, dry mouth also called xerostomia, problems swallowing also called dysphagia, tube feeding dependency (worst form of dysphagia) and hypothyroidism. All available NTCP models in head and neck cancer, assess their quality and compare predictive performance for predicting each type of side effect caused by radiotherapy were reviewed (Part 2 of this report: A literature review of NTCP models for head and neck cancer). These NTCP models can in individual patients estimate the potential benefit of the new proton radiation technique compared to photon radiation aiming at reduction of complications resulting from radiation.

- Task 1.4.1: A review on the effectiveness of proton therapy versus photon therapy for patients with head and neck cancer (UMCG leadership)

- Task 1.4.2: A literature review to identify the most suitable Normal Tissue Complication Probability (NTCP) models that can be used to describe the relationship between dose distributions and the risk of radiation-induced side effects observed after radiotherapy for head and neck cancer (for model-based approach) (UMCU Leadership)

These two reviews are also served as Phase 1 “Identification” and Sub-phase 1 “Learn from Past & Present” corresponding to the IHTAM framework, which is developed by HTx project team to aid to generalize the results of the case studies to other settings. Models identified in the Part 2 will be externally validated in existing available head and neck cancer datasets in Task 1.4.3, and methodological issues identified will be avoided in future model development.

# TASK 1.4.1: A REVIEW ON THE EFFECTIVENESS OF PROTON THERAPY VERSUS PHOTON THERAPY FOR PATIENTS WITH HEAD AND NECK CANCER.

## EXECUTIVE SUMMARY

**Background** Radiotherapy plays a pivotal role in the treatment of head and neck cancer patients. Next to the beneficial effects of radiotherapy on locoregional control and overall survival, head and neck radiotherapy may result in a broad spectrum of acute and late radiation-induced side effects, which have a major impact on quality of life of HNC survivors. Sine around 10 years, pencil beam scanning proton therapy (PBS-IMPT) has become commercially available, which is mainly applied to reduce the dose to the normal tissues, without jeopardising the therapeutic window by preventing radiation-induced side effect without affecting treatment efficacy. The objective of this literature review was to test to what extend proton therapy results in less toxicity without affecting efficacy in terms of locoregional control and survival.

**Methods** A systematic review was performed on studies that compared protons with photons with regard to acute and late radiation-induced toxicities, patient-rated outcome (PRO), and efficacy endpoints, including locoregional control, progression-free survival and overall survival. A PubMed search was performed to identify papers comparing protons with photons in squamous cell head and neck cancer originating in the oral cavity, larynx or pharynx. Pooled analyses were performed whenever possible using RevMan.

**Results** We identified 12 titles reporting on the direct comparison between protons and photons regarding the aforementioned endpoints. There were no RCT's available. Most studies were retrospective cohort studies of which many used prospectively collected data on toxicity and PRO's.

As PBS-IMPT is only available for HNC treatment since 5-10 years, most studies were relatively small with short follow up. Most studies found significant reductions of acute radiation-induced side effects, with a relative reduction of acute mucositis, dysphagia, xerostomia, dysgeusia and tube feeding dependence with relative reductions of approximately 50% or more. During the acute and subacute phase, significant reductions were seen regarding various PRO's.

No significant differences were observed regarding locoregional control, progression-free survival and/or overall survival. In nasopharyngeal cancer patients, a non-significant trend was observed towards better progression-free survival.

**Conclusion** At present, proton therapy in the treatment of HNC is mainly used to reduce the dose to the normal tissues with an equivalent dose to the target, aiming at reduction of radiation-induced side effects without jeopardising treatment efficacy. The results of the current literature review support that the dose reductions to organs-at-risk obtained with protons result in less acute

radiation-induced toxicities and patient-reported outcome, without jeopardising locoregional control, progression-free survival and overall survival.

## Introduction

Radiotherapy plays a pivotal role in the treatment of head and neck cancer (HNC). More than 50% of all HNC patients receive radiotherapy as their primary curative treatment modality, either alone or in combination with chemotherapy or targeted agents like cetuximab. Moreover, many patients treated with surgery require adjuvant radiotherapy, either alone or combined with chemotherapy, depending on the presence of adverse prognostic factors, like positive surgical margins and/or extranodal growth of lymph node metastases.

There is no doubt that radiotherapy heavily contributes to improving locoregional tumour control and overall survival. However, the downside is that radiotherapy induces both acute and late toxicity. Radiation-induced toxicity has a major impact on daily functioning and quality of life of HNC survivors [Langendijk 2008]. As overall survival of HNC patients significantly improved over the last decades, prevention of especially late radiation-induced toxicity becomes increasingly relevant.

As the risk of radiation-induced toxicity heavily depends on the dose in healthy tissues, new radiation technologies mainly aim at reducing the dose to the most relevant organs-at-risk. The most important radiation-induced toxicities after head and neck radiotherapy are xerostomia (dry mouth syndrome) and dysphagia. Some randomized controlled trials (RCT's) compared modern photon-based radiation techniques (e.g., 3D conformal radiotherapy (3D-CRT)) with modern photon-based techniques (e.g., intensity modulated radiotherapy (IMRT)) and showed that reducing the dose to the salivary glands resulted in significantly lower rate of late xerostomia [Kam 2007; Toledano 2012; Ghosh-Laskar 2016; Nutting 2011].

More recently, another new radiation technology has become commercially available, which is proton therapy. Due to its superior physical beam properties, the dose to many organs-at-risk can be markedly reduced and thus this technology is expected to widen the therapeutic ratio after definitive radiotherapy and chemoradiation in HNC patients by reducing the incidence of late radiation-induced toxicities without affecting locoregional tumour control and overall survival.

At present, results from RCT's comparing modern photon techniques, like IMRT and Volumetric Modulated Arc Therapy (VMAT), with modern proton techniques, like Intensity Modulated Proton Therapy (IMPT) are still ongoing, while the number of HNC patients treated with IMPT is rapidly rising.

Technology in proton therapy evolves rapidly. Approximately 10 years ago, the majority of centres used passive scattering proton therapy, which had limited benefit in HNC patients, as this technique was less suited for more complex target volumes with numerous organs-at-risk near the target. However, more recently, pencil beam scanning (PBS) proton therapy has become commercially available and virtually all proton therapy centres are currently using PBS proton therapy which allows for applying intensity modulated proton therapy (IMPT). Since the

introduction of PBS-IMPT in routine clinical practice, numerous planning comparative studies showed significant dose reductions in multiple organs-at-risk when applied in HNC as compared to IMRT and VMAT [Widesott 2008; Mathiessen 2008; Jakobi 2015; Swisher-McClure 2016].
It is generally assumed that, when the prescribed dose to the target (including the tumour and elective nodal areas) with photons is similar to that of photons, locoregional tumour control and subsequent survival should be similar as well. This assumption is supported by the findings of two RCT's comparing photons with protons in non-small cell lung cancer and esophageal cancer, respectively, in which locoregional control and overall survival were similar in both groups. However, the biological damage of proton therapy is generally considered to be higher than that of conventional photon radiotherapy using the same physical dose for tumours and healthy tissues [Paganetti 2002]. Therefore, current guidelines recommend a constant relative biological effectiveness (RBE) factor of 1.1 for clinical proton therapy treatments, meaning that the equivalent physical photon dose equals 1.1 times the physical proton dose which is also current routine clinical practice [ICRU 2007]. However, preclinical evidence suggests the RBE may vary with lower and higher RBE values depending on the location of tissues in relation to the Bragg-peak [Paganetti 2-14; Rorvic 2018; Wedenberg 2013], which theoretically could result in worse or better locoregional tumour control. Therefore, continuous attention should be paid to the possible effects of proton therapy on locoregional control in each tumour site.
At present, results from RCT's are currently not yet available. There are 3 RCT's ongoing in HNC and the first results are expected to become available in the next 3 to 5 years.
So far, PBS-IMPT is mainly used for reducing the dose to the most relevant organs-at-risk with a bioequivalent physical dose to the target, in order to broaden the therapeutic window by decreasing radiation-induced toxicity with equivalent efficacy in terms of locoregional tumour control and overall survival. As PBS-IMPT is clinically introduced on a larger scale during the last 5-10 years, the first results have only been published during the last 5 years.

*Objective*
Therefore, the aim to review the current literature to determine the efficacy of IMPT (protons) versus IMRT/VMAT (photons) in terms of locoregional tumour control and overall survival, and to evaluate if the dose reductions in organs-at-risk obtained with protons resulted in less radiation-induced toxicity and improved patient-rated outcome (PRO).

## Implementation

*Selection criteria*
We performed a literature review on papers published between 2016 and July 2021. This time frame was chosen as PBS-IMPT in HNC was only used during that period. Eligible were studies comparing state-of-the-art photon (IMRT or

VMAT) and proton techniques. To evaluate treatment outcome between protons and photons, we included studies which met the following criteria:

i) study design: all retrospective and prospective cohort and nested case-control studies;
ii) data source: studies that used routine care, registry data or data from randomized trials,
iii) studies that aimed to compare treatment outcome between photons and protons with regard to radiation-induced toxicity, PRO's and treatment efficacy (i.e., locoregional tumour control, overall survival, etcetera) in patient with head and neck cancer carcinoma originating in the oral cavity, oropharynx, hypopharynx, nasopharynx and larynx who were treated with curatively intended primary of postoperative radiotherapy.

*Search strategies*

A literature search was performed in MEDLINE. The following search terms were used: ("proton"[Title/Abstract] OR "Proton Therapy"[MeSH Terms]) AND ("radiotherap*"[Title/Abstract] OR "radiat*"[Title/Abstract] OR "Radiotherapy"[MeSH Terms]) AND ("head and neck"[Title/Abstract] OR "Head and Neck Neoplasms"[MeSH Terms] OR "oropharyn*"[Title/Abstract] OR "nasopharyn*"[Title/Abstract] OR "Oropharyngeal Neoplasms"[MeSH Terms] OR "Nasopharyngeal Neoplasms"[MeSH Terms] OR "Squamous Cell Carcinoma of Head and Neck"[MeSH Terms]).

Since this report served as an update on recent literature, only studies published from 2016 to August 2021 were considered for inclusion. Titles, abstracts and full-text articles were screened for relevance. Studies were included if they reported on radiation-induced toxicity, patient-rated outcome measures, locoregional control and/or progression-free survival after proton therapy versus photon therapy for HNC. Studies on oesophageal cancer, skull base tumours, paediatric cancer, rhabdomyosarcoma and sinonasal adenoid cystic carcinoma were excluded.

*Data collection and analysis*

After performing the search strategy described above, one review author selected the studies meeting the inclusion criteria. In case of doubt during the selection process, a second author was asked to review the abstracts. For all studies seemingly meeting the inclusion criteria based on title and/or abstract, the full text was reviewed. Special emphasis was placed on finding articles reporting on acute and late toxicity, PRO's and locoregional control and other efficacy endpoints.

Results for each study were reported based on the information available in the articles and appendices. Study characteristics were described for each study, including the number of participants, time period of treatment and details of radiotherapy.

*Data synthesis and meta-analysis approaches*

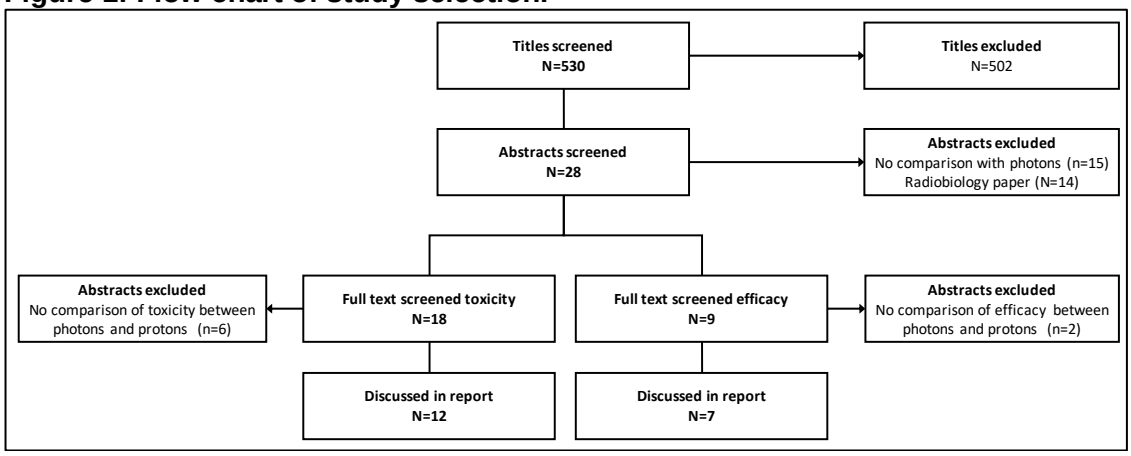
Methodology of the included studies, as well as results on radiation-induced toxicity, patient-rated outcome, progression-free survival, locoregional and distant control and overall survival, were reported. Additionally, progression-free

survival results of the selected studies on nasopharynx and oropharynx patients were pooled into a meta-analysis using a random effects model when sufficient data were available.
Salivary gland tumours were not included in the meta-analysis because of their differing biological behaviour. Log[Hazard Ratios] and Standard Errors of the log[Hazard Ratios] were extracted from all studies. When available, these numbers were obtained directly by calculating them from reported hazard ratios (HR) and 95% confidence intervals (CI). Otherwise, they were estimated based on number of events, numbers of patients in proton and photon therapy group, log rank P-value and direction of difference between the groups. The methods used for obtaining log[Hazard Ratios] and Standard Errors of the log[Hazard Ratios] are described in more detail by Tierny et al.(3) Results from the matched subgroups were used when available.
The program used for the meta-analysis was Review Manager (RevMan Version 5.4. The Cochrane Collaboration, 2020). Analyses were performed with all nasopharynx and oropharynx studies combined and separately for studies including either nasopharynx or oropharynx patients.

## Results

*Results of the search*
The selection process is depicted in **Figure 1**. The MEDLINE search resulted in 908 hits (as per 31 November 2021), of which 530 were studies published from 2016 onwards.

**Figure 1: Flow chart of study selection.**




Out of these, 28 titles were selected for review of abstracts. Eighteen abstracts on toxicity and PROMs were selected for screening of the full text of which twelve met the inclusion criteria.
In addition, nine titles were selected for screening of the full articles on efficacy outcome of which seven of these articles met all inclusion criteria that will be discussed in the present report.

*Description of studies*

The results of the selected studies are summarised in **Table 1**. As expected, no RCT's were found comparing photons with protons.

Four of the selected studies described efficacy endpoints for patients with nasopharyngeal cancer [Alterio 2020; Park 2019; Chou 2021; Li 2021) (**Table 1**). In two studies, the population consisted of oropharyngeal cancer patients [Yoon 2021; Blanchard 2016]. The remaining study included several types of head and neck cancer, mainly originating from the parotid or submandibular gland [Romesser 2016].

In all studies, modern state-of-the-art photon techniques were used, including IMRT, VMAT or helical tomotherapie. In almost all studies, PBT-IMPT was used, except in one in which uniform scanning was used in patients treated with unilateral irradiation [Romesser 2016]. In three studies, photon therapy was compared to a combination of photon therapy with proton therapy, in which generally only the boost was given with proton therapy [Yoon 2021; Alterio 2020; Park 2019]. It should be stressed that in these studies, the full potential of protons was not explored as only a part of the fractions was administered with protons. In general, fractionation schedules and prescribed target dosages used were similar between photon and protons in all studies.

Most investigators used some kind of matching technique to account for confounding, including propensity score matching [Chou 2021; Li 2021; Yoon 2021; Park 2019; Sio 2016], or matched-case analysis [Yasuda 2021; Blanchard 2016; Holliday 2015].

Ten out of 12 studies reported on physician-rated acute toxicity (**Table 1**), while late toxicity was only reported in 2 studies [Alterio 2020; Li 2021] and PRO's in only 3 studies [Manzar 2020; Sio 2016; Cao 2021].

**Table 1: Summary of study characteristics**

| Reference | Tumour site | Radiotherapy | | Study design | Numbers | Modalities | Endpoints | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Photons | Protons | | | | Acute toxicity | Late toxicity | Hospital admission | PRO | LRC | PFS | OS |
| **Cao 2021** | Oropharynx | IMRT-SIB with total dose of 54-63/66-70 Gy | IMPT-SIB with total dose of 54-63/66-70 Gy | Retrospective comparison with prospective patient-rated xerostomia scoring | PhT: 429 patients PrT: 103 patients | Concurrent CRT or induction CHT + RT | No | No | No | YES | No | No | No |
| **Yasuda 2021** | Pharynx | Sequential boost IMRT with 2 dose levels: 46/70 Gy | Sequential boost IMPT with 2 dose levels: 46/70 Gy | Retrospective comparison + model-based clinical evaluation + matched pair analysis | PhT: 127 patients PrT: 15 patients | Concurrent CRT +/+ induction CHT or concurrent CRT + adjuvant RT | YES | No | No | No | No | No | No |
| **Chou 2021** | Nasopharynx | IMRT-SIB with 3 dose levels of 54.12/59.4/ 69.96 Gy | IMPT-SIB with 3 dose levels of 54.12/59.4/ 69.96 Gy | Propensity score-matched analysis | PhT: 80 patients PrT: 80 patients | Concurrent CRT or induction CHT + RT or RT alone | YES | No | Yes | No | No | YES | YES |
| **Li 2021** | Nasopharynx | IMRT-SIB with 3 dose levels of 54.12/59.4/69.96 Gy or 56/59-63/70 Gy | IMPT-SIB with 3 dose levels of 54.12/59.4/69.96 Gy or 56/59-63/70 Gy | Propensity score-matched analysis with retrospective toxicity scoring | PhT: 49 patients PrT: 28 patients 2 x 24 patients used for PSM | Concurrent RT or RT alone | YES | YES | No | No | YES | YES | YES |
| **Yoon 2021** | Oropharynx | IMRT-SIB with 3 dose levels of 36.0/60.0/68.4 Gy (30 fx) | Mixed beam with IMRT-SIB with 3 dose levels of 36.0/60.0/68.4 Gy (30 fx) with last 12 fx using IMPT | Propensity score-matched analysis | PhT: 81 patients PrT: 67 patients 2 x 36 patients used for PSM | Concurrent CRT | YES | No | No | No | No | YES | YES |
| **Alterio 2020** | Nasopharynx | IMRT only with 3 dose levels SIB-technique (70.0/59.4/56.1 Gy | Mixed beam with IMRT (54 Gy) + PrT boost to 70-74 Gy | Historical comparison with prospective toxicity scoring for PrT and retrospective toxicity scoring for PhT | PhT: 17 patients PrT: 27 patients | Definitive RT with CRT and only 1 with RT alone | YES | YES | No | No | YES | No | No |
| **Manzar 2020** | Oropharynx | VMAT to total dose of 60-66 Gy (small volume GTV) or 70 Gy (large volume) | IMPT to total dose of 60-66 Gy (small volume GTV) or 70 Gy (large volume) | Retrospective comparison woth prospective toxicity scoring and PROM's | PhT: 46 patients PrT: 259 patients | Concurrent CRT or RT alone | YES | No | YES | YES | No | No | No |
| **Park 2019** | Nasopharynx | Helical Tomo-SIB with 3 dose levels of 36.0/60.0/68.4 Gy (30 fx) | Mixed beam with helical Tomo-SIB with 3 dose levels of 36.0/60.0/68.4 Gy (30 fx) with last 12 fx using IMPT | Propensity score-matched analysis | PhT: 63 patients PrT/PhT: 35 patients 2 x 35 patients used for PSM | Concurrent CRT | YES | No | No | No | No | YES | No |
| **Blanchard 2016** | Oropharynx | IMRT-SIB with total dose of 54-63/66-70 Gy | IMPT-SIB with total dose of 54-63/66-70 Gy | Matched-case analysis | PhT: 100 patients PrT: 50 patients | Concurrent CRT or induction CHT + RT | YES | No | No | No | YES | YES | YES |
| **Sio 2016** | Oropharynx | IMRT-SIB with 3 dose levels of 57/63/70 Gy | IMPT-SIB with 3 dose levels of 57/63/70 Gy | Retrospective comparison with prospective PRO scoring | PhT: 46 patients PrT: 35 patients | Concurrent CRT or induction CHT + RT | No | No | No | YES | No | No | No |
| **Romesser 2016** | Salivary gland and skin cancer | Unilateral IMRT to total dose 60/66/70 Gy depending on risk factors | Uniform scanning PrT to total dose 60/66/70 Gy depending on risk factors | Retrospective comparison with prospective toxicity scoring | PhT: 23 patients PrT: 18 patients | Postoperative RT or CRT | YES | No | No | No | YES | No | YES |
| **Holliday 2015** | Nasopharynx | IMRT-SIB with 3 dose levels of 57/63/70 Gy | IMPT-SIB with 3 dose levels of 57/63/70 Gy | Matched-case analysis with prospective toxicity scoring of acte toxicity | PhT: 20 patients PrT: 10 patients | Concurrent CRT +/- induction CHT or BioRT of RT alone | YES | No | No | No | No | No | No |

BioRT = bioradiation = radiotherapy + cetuximab
CRT = Chemoradiation
IMRT-SIB = Intensity Modulated Radiotherapy with Simultaneous Integrated Boost
LRC = Locoregional Control
PhT = Photon therapy
OS = Overall survival
PFS = Progression-free survival
PRO = Patient-rated Outcome
PrT = Proton therapy
PSM = Propensity score analysis
RT = Radiotherapy

*Patient-rated outcome*

In four studies, PRO's were prospectively scored both among patients treated with protons and photons as part of a prospective data registration program (**Table 2**) [Sio 2016; Manzar 2020; Blanchard 2016; Cao 2016]. These studies only included patients with oropharyngeal cancer. In one study [Manzar 2020], different subsets of patients were investigated, including those treated with definitive radiotherapy, bilateral radiotherapy, unilateral radiotherapy, chemoradiation and RT alone. Various instruments were used to assess PRO's, including the EORTC QLQ-H&N35, the XQ questionnaire and the MDASI.

**Table 2: Overview of significant differences found for patient-rated outcome in the different studies. Green cells indicate significantly better outcome in favour of proton therapy, while the orange cells indicate worse outcome for protons compared to photons. The grey cells indicate no difference between photons and protons.**

| | Sio 2016 | Manzar 2020 | Manzar 2020 | Manzar 2020 | Manzar 2020 | Manzar 2020 | Blanchard 2016 | Cao 2021 |
|---|---|---|---|---|---|---|---|---|
| Setting | CRT | Definitive RT | Bilateral RT | Unilateral RT | CRT | RT alone | CRT | Definitive RT of CRT |
| RT techniques compared | IMRT vs. IMPT | VMAT vs. IMPT | VMAT vs. IMPT | VMAT vs. IMPT | VMAT vs. IMPT | VMAT vs. IMPT | IMRT vs. IMPT | IMRT vs. IMPT |
| Number of patients | 46 vs. 35 | 111 vs. 27 | 221 vs. 40 | 38 vs. 6 | 173 vs. 36 | 86 vs. 10 | 100 vs. 50 | 429 vs. 103 |
| Tumour site | Oropharynx | Oropharynx | Oropharynx | Oropharynx | Oropharynx | Oropharynx | Oropharynx | Oropharynx |
| Study design | Prospective | Prospective | Prospective | Prospective | Prospective | Prospective | Mtached pair analysis | Cross-sectional |
| Cough (EORTC QLQ H&N35) | | | p=0.011 | | | p=0.0002 | | |
| Sense of smell/taste (EORTC QLQ H&N35) | | | p=0.053 | | p=0.034 | p=0.021 | | |
| Problems with teeth (EORTC QLQ H&N35) | | p=0.028 | | p=0.048 | | p=0.049 | | |
| Dry mouth (EORTC QLQ H&N35) | | | | p=0.039 | | | | |
| Nutritional supplements (EORTC QLQ H&N35) | | p=0.017 | p=0.032 | | p=0.009 | | | |
| Swallow (EORTC QLQ H&N35) | | p=0.058 | | | p=0.009 | p=0.024 | | |
| Feeling ill (EORTC QLQ H&N35) | | p=0.026 | | | p=0.025 | | | |
| Sexual symptoms (EORTC QLQ H&N35) | | | | | | 0.0875 | | |
| Feeding tube (EORTC QLQ H&N35) | | p=0.027 | | | | | | |
| Sticky saliva (EORTC QLQ H&N35) | | | | p=0.044 | | | | |
| Weight loss (EORTC QLQ H&N35) | | | | | | p=0.068 | | |
| Patient-rated xerostomia 3 months after RT | | | | | | | p=0.009 | |
| Patient-rated xerostomia 1 year after RT | | | | | | | p=0.23 | |
| Patient-rated fatigue 3 months after RT | | | | | | | p=0.80 | |
| Patient-rated fatigue 1 year after RT | | | | | | | p=0.17 | |
| Xerostomia (XQ--questionnaire) 0-3 months | | | | | | | | p=0.50 |
| Xerostomia (XQ--questionnaire) 3-6 months | | | | | | | | p=1.00 |
| Xerostomia (XQ--questionnaire) 6-9 months | | | | | | | | p=0.85 |
| Xerostomia (XQ--questionnaire) 9-12 months | | | | | | | | p=0.61 |
| Xerostomia (XQ--questionnaire) 12-18 months | | | | | | | | p=0.24 |
| Xerostomia (XQ--questionnaire) 18-24 months | | | | | | | | p=0.025 |
| Xerostomia (XQ--questionnaire) 24-36 months | | | | | | | | p=0.010 |
| Food taste (MDASI) acute phase | ns | | | | | | | |
| Dry mouth (MDASI) acute phase | ns | | | | | | | |
| Swallowing/chewing (MDASI) acute phase | ns | | | | | | | |
| Fatigue (MDASI) acute phase | ns | | | | | | | |
| Pain (MDASI) acute phase | ns | | | | | | | |
| Appetite (MDASI) acute phase | ns | | | | | | | |
| Mucus (MDASI) acute phase | ns | | | | | | | |
| Sleep (MDASI) acute phase | ns | | | | | | | |
| Mouth sores (MDASI) acute phase | ns | | | | | | | |
| Drowsiness (MDASI) acute phase | ns | | | | | | | |
| Distress (MDASI) acute phase | ns | | | | | | | |
| Food taste (MDASI) 3 months | p=0.003 | | | | | | | |
| Dry mouth (MDASI) 3 months | ns | | | | | | | |
| Swallowing/chewing (MDASI) 3 months | ns | | | | | | | |
| Fatigue (MDASI) 3 months | ns | | | | | | | |
| Pain (MDASI) 3 months | ns | | | | | | | |
| Appetite (MDASI) 3 months | ns | | | | | | | |
| Mucus (MDASI) 3 months | p=0.038 | | | | | | | |
| Sleep (MDASI) 3 months | ns | | | | | | | |
| Mouth sores (MDASI) 3 months | ns | | | | | | | |
| Drowsiness (MDASI) 3 months | ns | | | | | | | |
| Distress (MDASI) 3 months | ns | | | | | | | |
| Dry mouth (MDASI) chronic phase | ns | | | | | | | |
| Swallowing/chewing (MDASI) chronic phase | ns | | | | | | | |
| Fatigue (MDASI) chronic phase | ns | | | | | | | |
| Pain (MDASI) chronic phase | ns | | | | | | | |
| Appetite (MDASI) chronic phase | ns | | | | | | | |
| Mucus (MDASI) chronic phase | ns | | | | | | | |
| Sleep (MDASI) chronic phase | ns | | | | | | | |
| Mouth sores (MDASI) chronic phase | ns | | | | | | | |
| Drowsiness (MDASI) chronic phase | ns | | | | | | | |
| Distress (MDASI) chronic phase | ns | | | | | | | |

In summary, in some studies, significant benefits of protons compared to photons were found for patient-rated complaints most of them related to cough, sense of taste and smell, problems with teeth, xerostomia, use of nutritional supplements, and weight loss, while conflicting results were found for dysphagia (**Table 2**).
Most studies only reported on PRO's during the course of radiation or immediately after completion of treatment. One study reported on xerostomia up to 24 months after completion of treatment (**Table 2**) [Cao 2021]. They found no difference in patient-reported xerostomia in the first 12 months after completion of treatment, but a significant recovery at 18 and 24 months, respectively, among those treated with proton therapy.
Unfortunately, the available data did not allow for a pooled analysis on the effects of PRO's.

*Acute toxicities*

Ten studies reported on acute toxicity that occurred during the course of radiation up to 3 months after completion of treatment (**Table 3**). In the majority of

endpoints investigated in the different studies, physician-rated scores for acute toxicity were significantly lower among patients treated with protons. This was particularly true for oral mucositis, dysgeusia, toxicity endpoints directly or indirectly related to swallowing disorders (i.e., dysphagia, tube feeding dependence, anorexia and different types of weight loss) and pain (i.e., pain, oral pain, pharyngeal pain and use of narcotics required at the end of treatment). Acute dermatitis was significantly more severe among patients treated with protons.

In one study with a relatively high number of patients [Manzar 2020], different subsets of patients were distinguished to see if the differences regarding acute toxicity was different depending on treatment modality (**Table 4**). The differences between protons and photons were most pronounced among those treated with postoperative radiotherapy and those treated with concurrent chemoradiation. The most important acute toxicity during the course of radiation is acute mucositis, leading to a number of subsequent side effects and complaints like pain, often requiring narcotics, dysphagia and subsequent tube feeding dependence and weight loss. Seven studies reported on acute mucositis grade $\geq 3$, including a total number of 262 patients treated with protons and 492 patients treated with photons (**Figure 2**). The overall effect was statistically significant in favour of protons with an RR of 0.40 (95%-ci: 0.28-0.58; $p<0.001$), which corresponds to an absolute reduction of 16% (95%-ci: 6% - 46%). No significant heterogeneity across studies was observed.

As in a number of studies, similar or comparable endpoints were used which enabled a pooled analysis of results.

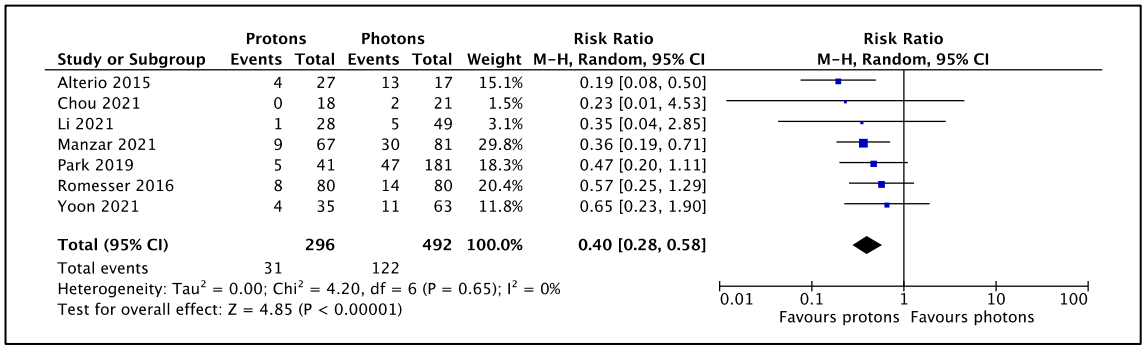
**Table 3: Overview of physician-rated acute toxicities of photons versus protons. Green cells indicate significantly better outcome in favour of proton therapy, while the orange cells indicate worse outcome for protons compared to photons. The grey cells indicate no difference between photons and protons.**

| | Li 2021 | Manzar 2020 | Romesser 2016 | Chou 2021 | Alterio 2020 | Yasuda 2021 | Holliday 2015 | Yoon 2021 | Blanchard 2016 | Park 2019 |
|---|---|---|---|---|---|---|---|---|---|---|
| Setting | Definitive RT | ALL | Unilateral POSTOP RT | Definitive RT | CRT | CRT | Definitive RT | CRT | Bilateral RT | CRT |
| RT techniques compared | IMRT vs. IMPT | VMAT vs. IMPT | IMRT vs. Uniform scanning | VMAT vs. IMPT | IMRT vs. Mixed Beam | IMRT vs. IMPT | IMRT vs. IMPT | IMRT vs. Mixed Beam | IMRT vs. IMPT | TOMO vs. Mixed Beam |
| Number of patients | 49 vs. 28 | 259 vs. 46 | 23 vs. 18 | 80 vs. 80 | 27 vs. 17 | 108 vs. 15 | 20 vs. 10 | 36 vs. 36 | 100 vs. 50 | 35 vs. 35 |
| Tumour site | Nasopharynx | Oropharynx | Salivary gland | Nasopharynx | Nasopharynx | Mixed pharynx | Nasopharynx | Oropharynx | Oropharynx | Nasopharynx |
| Study design | Prospective proponsity score analysis | Prospective | Retrospective | Retrospective propensity score analysis | Retrospective historical comparison | | Prospective matched control study | Retrospective propensity score analysis | Mtached pair analysis | Prospective propensity score analysis |
| Oral mucositis | p=0.03 | p=0.004 | p=0.005 | p=0.178 | p=0.0002 | | | p=0.012 | p=0.90 | p=0.358 |
| G-tube needed ≤ 30 days | | p=0.001 | | p=0.026 | p=0.81 | | p=0.020 | | | |
| Dysphagia | p=0.05 | p=0.073 | p=0.101 | | p=0.36 | p=0.0115 | p=0.175 | | | |
| Pain | | p=0.0004 | | | p=0.34 | | | | | |
| Oral pain | p=0.66 | p=0.0085 | | | | | | | | |
| Fatigue | p=0.02 | | p=0.002 | | | | | | p=0.13 | |
| Acute hospitalizations ≤ 60 days | | p=0.009 | | | | | | | | |
| Narcotic required at the end of RT | | p=0.017 | | | | | | | | |
| Weight loss | p<0.001 | p=0.06 | | | p=0.11 | | | p=0.071 | | p=0.484 |
| Dry mouth | p=0.002 | | | p=0.358 | p=0.02 | p=0.4273 | | | | |
| Nausea | p=0.03 | | p=0.003 | | | | | | | |
| Mean weight loss | | p=0.1 | | p=0.038 | | | | | | |
| Average morfine equivalent | | p=0.038 | | | | | | | | |
| Pharyngeal pain | | | | | | | | | | |
| Dysgeusia | p=0.004 | | p=<0.001 | | p=0.55 | p=0.0261 | | | | |
| Anorexia | | p=0.0695 | | | | | | | | |
| Acute grade ≥ 3 toxicities | | | | | | | p=0.015 | | | |
| Dehydration | | | | | | | | | | |
| Hoarseness | p=0.007 | | | | | | | | | |
| Weight loss > 7% during RT | | | | p=0.006 | | | | | | |
| G-tube at 3 months post RT | | | | | | | | | p=0.10 | |
| Analgesic use | | | | | | | | p=0.085 | | p=0.382 |
| Weight loss > 7% during RT or G-tube needed | | | | p=0.002 | | | | | | |
| Dysphonia | | | | | p=0.06 | | | | | |
| Hearing impairment | | | | | p=0.64 | | | | | |
| Grade 3 weight loss | | | | | | | | | | |
| Weight loss grade ≥ 3 during RT | | | | | | | | | p=0.11 | |
| Weight loss before G-tube during RT | | | | | | | p=0.333 | | | |
| Median duration of PEG-tube | | | | | | | p=0.230 | | p=0.12 | |
| Unschedeled hospitalization | | | | p=0.786 | | | | | p=0.84 | |
| Emergency room visit | | | | p=0.151 | | | | | p=0.89 | |
| Laryngeal edema | | | | | | | | | | |
| Duct inflammation | | | | | | | | | | |
| Mucosal infections | | p=0.015 | | | | | | | | |
| Dermatitis | p=0.45 | p=0.073 | p=0.032 | p<0.001 | p=0.66 | | p=0.049 | p=0.235 | p=0.15 | p=0.453 |

**Table 4: Overview of physician-rated acute toxicities of photons versus protons in the different subsets of the study of Manzar 2020. Green cells indicate significantly better outcome in favour of proton therapy, while the orange cells indicate worse outcome for protons compared to photons. The grey cells indicate no difference between photons and protons.**

| | Manzar 2020 | Manzar 2020 | Manzar 2020 | Manzar 2020 | Manzar 2020 | Manzar 2020 | Manzar 2020 |
|---|---|---|---|---|---|---|---|
| Setting | POSTOP RT | ALL | CRT | RT alone | Bilateral RT | Unilateral RT | Definitive RT |
| RT techniques compared | VMAT vs. IMPT | VMAT vs. IMPT | VMAT vs. IMPT | VMAT vs. IMPT | VMAT vs. IMPT | VMAT vs. IMPT | VMAT vs. IMPT |
| Number of patients | 148 vs.19 | 259 vs. 46 | 173 vs. 36 | 86 vs. 10 | 221 vs. 40 | 38 vs. 6 | 111 vs. 27 |
| Tumour site | Oropharynx | Oropharynx | Oropharynx | Oropharynx | Oropharynx | Oropharynx | Oropharynx |
| Study design | Prospective | Prospective | Prospective | Prospective | Prospective | Prospective | Prospective |
| Oral mucositis | p=0.0001 | p=0.004 | p=0.011 | p=0.003 | p=0.027 | p=0.0011 | p=0.056 |
| G-tube needed ≤ 30 days | p=0.069 | p=0.001 | p<0.0001 | p=0.099 | p=0.004 | p=0.083 | p<0.0001 |
| Dysphagia | | p=0.073 | p=0.0016 | p=0.015 | p=0.051 | p=0.003 | |
| Pain | p=0.0002 | p=0.0004 | p=0.0014 | | p=0.006 | p=0.035 | p=0.082 |
| Oral pain | p=0.034 | p=0.0085 | | p=0.0001 | p=0.06 | p=0.003 | |
| Fatigue | p=0.01 | | | p=0.0001 | | p=0.058 | |
| Acute hospitalizations ≤ 60 days | | p=0.009 | p=0.002 | | p=0.002 | | p<0.0001 |
| Narcotic required at the end of RT | p=0.0055 | p=0.017 | p=0.009 | | p=0.02 | | |
| Weight loss | p=0.018 | p=0.06 | p=0.074 | | p=0.081 | p=0.011 | |
| Dry mouth | | | | p=0.029 | p=0.079 | | |
| Nausea | p=0.028 | | | | | | |
| Mean weight loss | p=0.037 | p=0.1 | p>0.05 | p>0.05 | p>0.05 | p>0.05 | p>0.05 |
| Average morfine equivalent | p=0.03 | p=0.038 | p=0.057 | | | | |
| Pharyngeal pain | p=0.043 | | | p=0.003 | | | |
| Dysgeusia | | | p=0.04 | | | | |
| Anorexia | p=0.016 | p=0.0695 | p=0.078 | | | | |
| Acute grade ≥ 3 toxicities | | | | | | | |
| Dehydration | p=0.034 | | | | | | |
| Hoarseness | | | | | | | |
| Weight loss > 7% during RT | | | | | | | |
| G-tube at 3 months post RT | | | | | | | |
| Analgesic use | | | | | | | |
| Weight loss > 7% during RT or G-tube needed | | | | | | | |
| Dysphonia | | | | | | | |
| Hearing impairment | | | | | | | |
| Grade 3 weight loss | | | | | | | |
| Weight loss grade ≥ 3 during RT | | | | | | | |
| Weight loss before G-tube during RT | | | | | | | |
| Median duration of PEG-tube | | | | | | | |
| Unschedeled hospitalization | | | | | | | |
| Emergency room visit | | | | | | | |
| Laryngeal edema | | | p=0.053 | p=0.025 | | | p=0.075 |
| Mucosal infections | | p=0.015 | | p=0.07 | | | |
| Duct inflammation | | | | p=0.004 | | p=0.036 | |
| Dermatitis | | p=0.073 | | p=0.057 | p=0.024 | p=0.029 | p=0.1 |

**Figure 2: Pooled analysis of protons versus photons regarding ACUTE MUCOSITIS grade ≥ 3.**




Dysphagia is one of the most frequently reported toxicities during the course of radiation which has a major impact on quality of life [Langendijk 2008]. Four studies reported on dysphagia grade ≥ 2 (only able to eat soft food or worse), including a total number of 87 patients treated with protons and 197 patients treated with photons [Romesser 2016; Alterio 2020; Yasuda 2021; Li 2021]. The overall effect was statistically significant in favour of protons (**Figure 3**), with an RR of 0.42 (95%-ci: 0.24-0.75; p=0.004), which corresponds to an absolute risk reduction of 17% (95% CI: 7% - 42%). No significant heterogeneity across studies was observed.

**Figure 3: Pooled analysis of protons versus photons regarding DYSPHAGIA grade ≥ 2.**

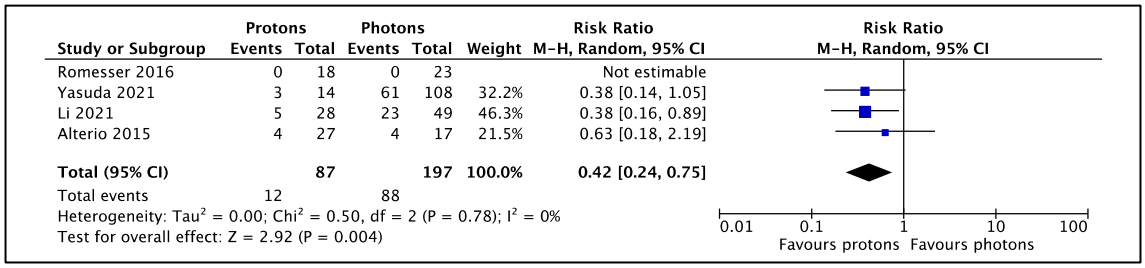
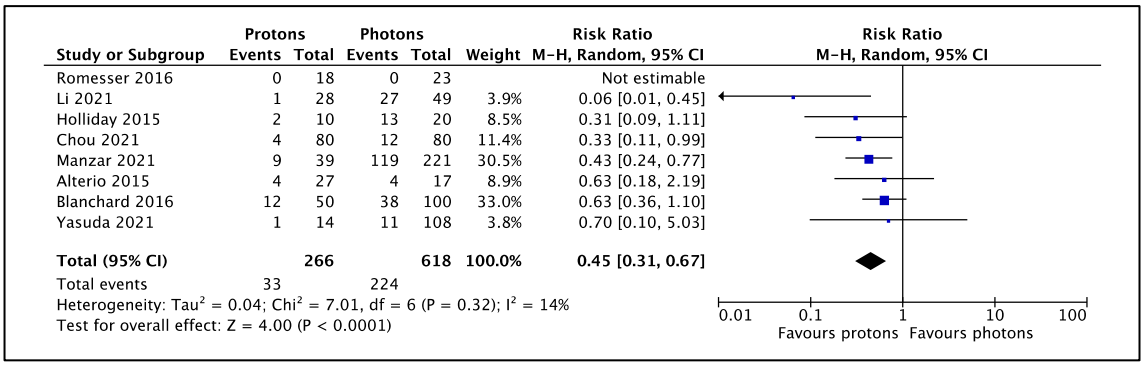
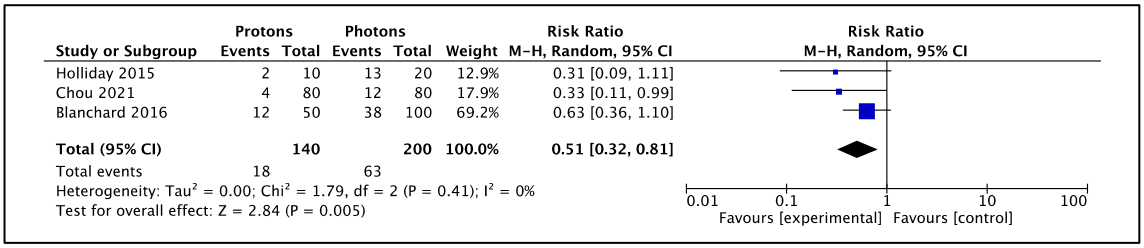


**Figure 4: Pooled analysis of all studies with available data comparing protons versus photons regarding DYSPHAGIA grade ≥ 3 (A) and the subset of studies including only studies that used a matching technique like propensity score matching or case-control matching (B).**

**A**




**B**

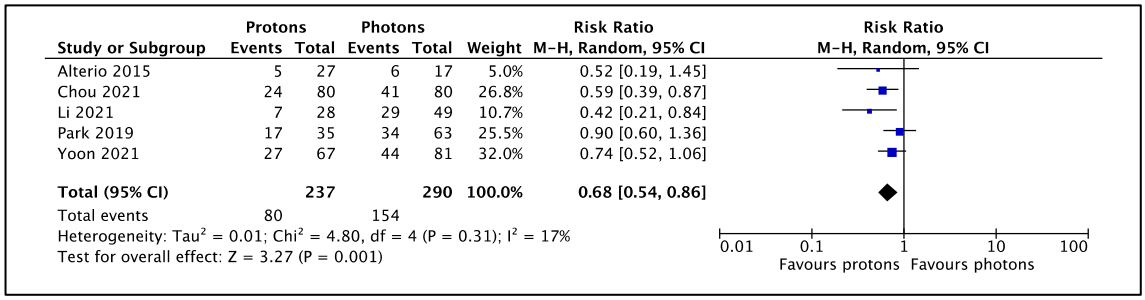



Four studies reported on dysphagia grade ≥ 3 (i.e., Only able to take liquid food or tube feeding dependent), including a total number of 266 patients treated with protons and 618 patients treated with photons (**Figure 4A**). The overall effect was statistically significant in favour of protons with an RR of 0.45 (95%-ci: 0.31-0.67; p<0.0001), which corresponds to an absolute reduction of reduction of 19% (95%-ci: 5% - 32%). No significant heterogeneity across studies was observed. For this endpoint, sufficient data was available to perform a pooled subset analysis on the studies that used some kind of matching technique, like matched case control or propensity score analysis (**Figure 4B**). In this analysis, the RR was approximately similar: RR = 0.51 (85%-ci: 0.32-0.81; p<0.005).

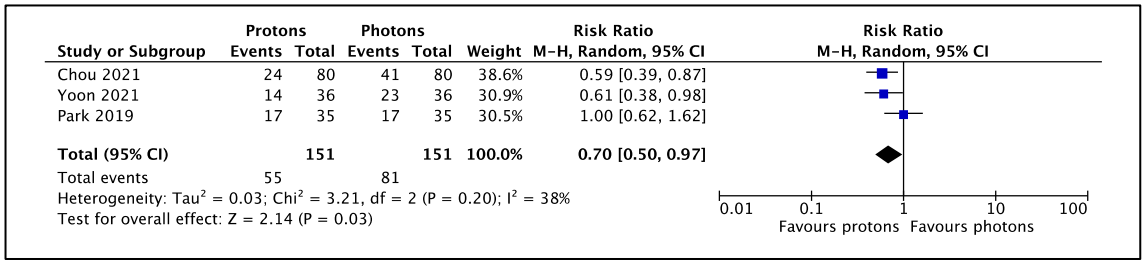
Weight loss results from acute mucositis, leading to pain particularly with intake of food and thus to dysphagia and subsequent weight loss. Five studies reported on weight loss during the course of radiation, including 237 patients treated with protons and 290 patients treated with photons (**Figure 5A**). The overall effect was statistically significant in favour of protons with an RR of 0.68 (95%-ci: 0.54-0.86; p<0.0001), which corresponds to an absolute reduction of reduction of 18% (95%-ci: 10% - 27%). No significant heterogeneity across studies was observed. For this endpoint, sufficient data was available to perform a pooled subset analysis on the studies that used some kind of matching technique, like matched case control or propensity score analysis (**Figure 5B**). In this analysis, the RR was approximately similar: RR = 0.70 (85%-ci: 0.50-0.97; p=0.03).

**Figure 5: Pooled analysis of all studies with available data on WEIGHT LOSS grade ≥ 2 comparing photons with protons (A) and including only studies that used a matching technique like propensity score matching or case-control matching (B).**

**A**

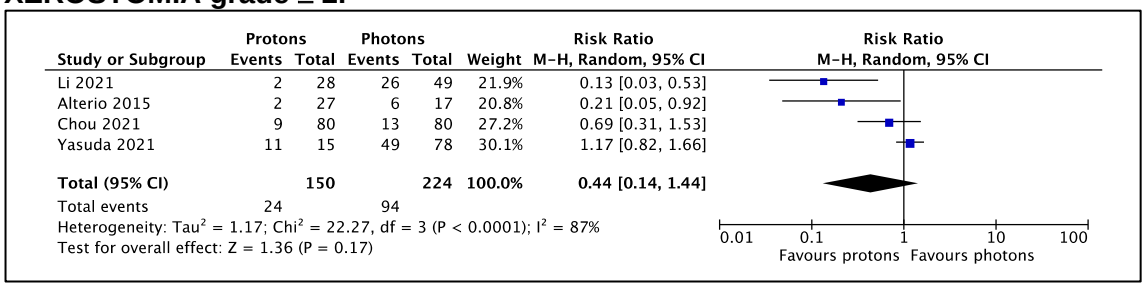



**B**



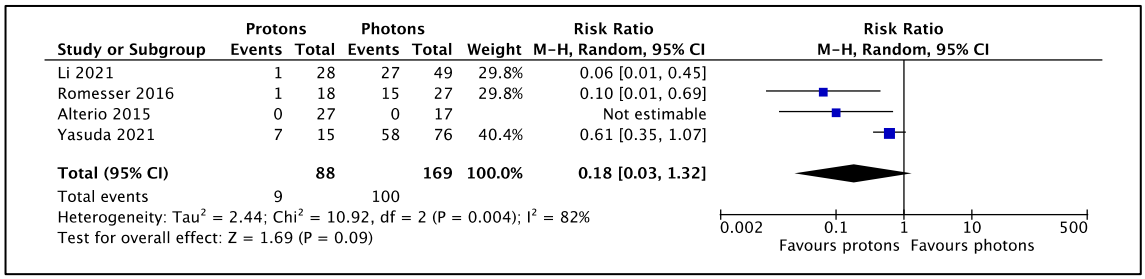

Acute xerostomia is a frequently reported acute toxicity already occurring in the first 3 weeks during the radiation course. Four studies reported on acute xerostomia grade ≥ 2, including a total number of 150 patients treated with protons and 224 patients treated with photons (**Figure 6**). The overall effect was not statistically significant with a RR of 0.44 (95%-ci: 0.14-1.44; p=0.159) in favour of protons, which corresponds to an absolute reduction of reduction of 17% (95%-ci: 7% - 44%). For this endpoint, significant heterogeneity across studies was observed (p< 0.0001).

**Figure 6: Pooled analysis of protons versus photons regarding ACUTE XEROSTOMIA grade ≥ 2.**




Four studies reported on dysgeusia grade ≥ 2, including a total number of 88 patients treated with protons and 169 patients treated with photons (**Figure 7**). The RR of the overall effect 0.18 (95%-ci: 0.02-2.95; p<0.001), which corresponds to an absolute risk reduction of 25% (95%-ci: 10% - 60%) in favour of protons. However, this difference was not statistically significant (p=0.09).

**Figure 7: Pooled analysis of protons versus photons regarding DYSGEUSIA grade ≥ 2.**

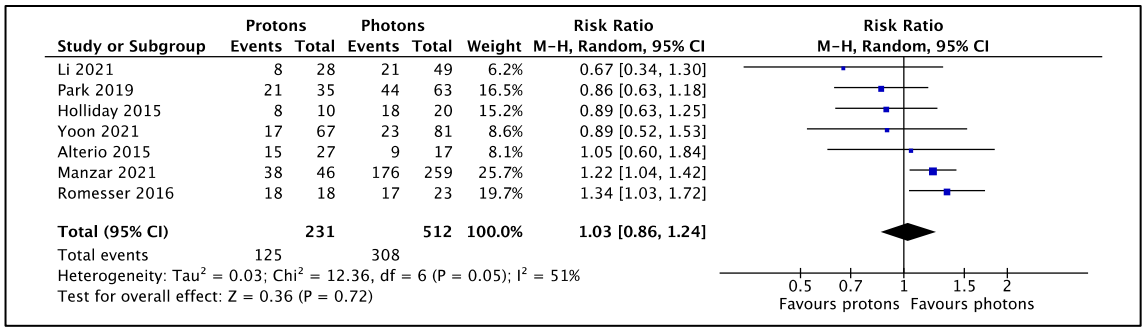



Seven studies reported on acute dermatitis grade ≥ 2, including a total number of 231 patients treated with protons and 512 patients treated with photons (**Figure 8A**). The overall effect was not statistically significant an RR of 1.03 (95%-ci: 0.86-1.24; p=0.72).However, significant heterogeneity was observed ($I^2$= 50%; p=0.05).

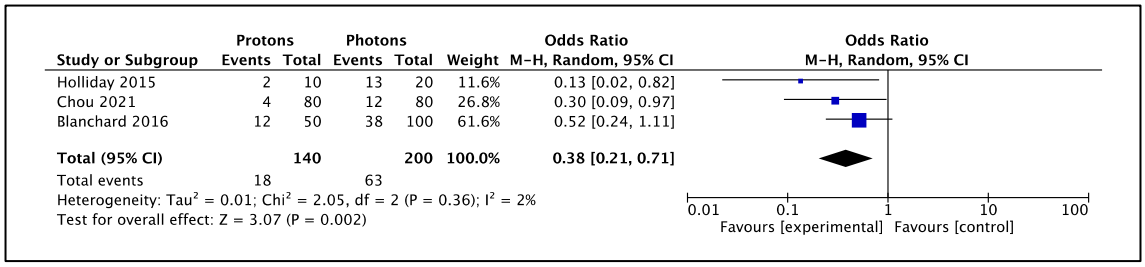
In 3 studies, a kind of matching was performed (**Figure 8B**), which showed significantly less acute dermatitis grade ≥ 2 among patients treated with protons, with an RR of 0.38 (95%-ci: 0.21 0.71; p=0.002)

**Figure 8: Pooled analysis of all studies with available data on ACURE DERMATITIS grade ≥ 2 comparing photons with protons (A) and including only studies that used a matching technique like propensity score matching or case-control matching (B).**

**A**




**B**

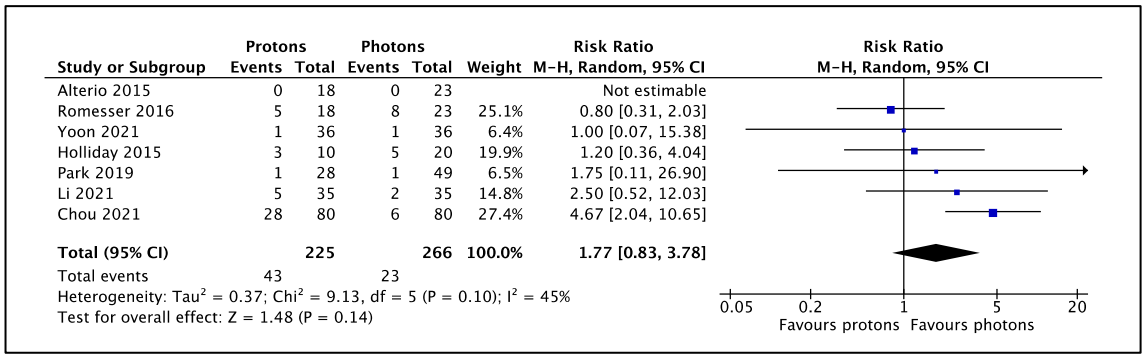



Seven studies reported on acute dermatitis grade ≥ 3, including a total number of 225 patients treated with protons and 266 patients treated with photons (**Figure 9A**). The overall effect was not statistically significant with an OR of 1.77 (95%-ci: 0.83-3.78; p=0.14).

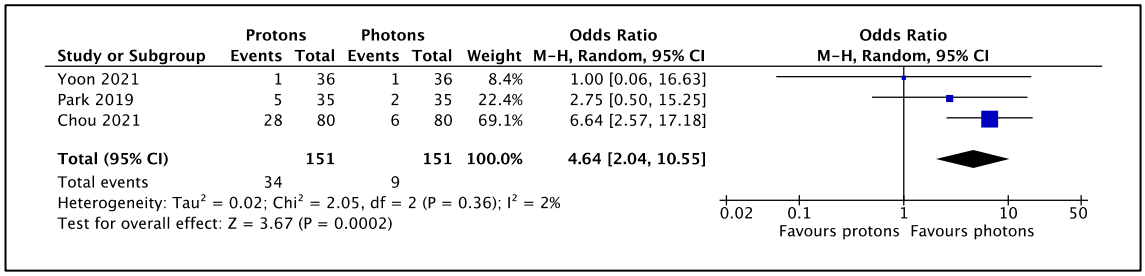
In 3 studies, a kind of matching was performed (**Figure 9B**), which showed significantly more acute dermatitis grade ≥ 3 among patients treated with protons, with an RR of 4.64 (95%-ci: 2.04 - 10.55; p=0.0002).

**Figure 9: Pooled analysis of all studies with available data on ACURE DERMATITIS grade ≥ 3 comparing photons with protons (A) and including only studies that used a matching technique like propensity score matching or case-control matching (B).**

**A**




**B**




*Late toxicities*

As mentioned in the introduction, PBS-IMPT is a relatively new technique and consequently, the number of patients with long-term follow up is limited. Reports on late toxicity are mainly published by one centre (i.e., MD Anderson Cancer Center), that has the longest experience in treating HNC patients with proton therapy [Blanchard 2016; Cao 2021; Sio 2016, Zhang 2017].

Cao, et al reported on the largest series of oropharyngeal cancer patients treated with either IMRT (n=429) or IMPT (n=103) [Cao 2021].In this study, the authors focussed on the development of patient-rated xerostomia from the start of radiotherapy up to 36 months after completion of treatment. In the first 18 months, no differences were noted regarding patient-rated xerostomia. However, patients treated with IMPT reported significantly lower rates of moderate-to-severe xerostomia at 24 (20% vs. 6%; p=0.025) to 36 months (20% vs. 6%; p=0.01). Xerostomia at later time points (24 and 36 months) was significantly associated with the dose to the oral cavity, which was significantly lower with IMPT). Strikingly, salivary gland dose was not associated with late patient-reported xerostomia.

Blanchard, at al. performed a matched case control study consisting of 50 oropharyngeal cancer patients treated with IMPT and 100 oropharyngeal cancer patients treated with IMRT, which all received concurrent chemotherapy as well. At 3 months and 1 year after treatment, patients treated with IMPT showed lower rates of the pre-planned composite endpoint of grade ≥ 3 weight loss or tube feeding dependence, with an OR of 0.44 (95%-ci: 0.19-1.00; p=0.05) and 0.23 (95%-ci: 0.07-0.73) [Blanchard 2016].
Zhang, et al. reported on the incidence of osteoradionecrosis (ORN) among oropharyngeal cancer patients treated with IMPT versus IMRT, which is one of the most severe side effects after head and neck radiotherapy [Zhang 2017]. The risk of ORN is associated with the dose to the mandible, which could be reduced significantly with IMPT as compared to with IMRT (mean dose: 25.6 Gy vs. 41.2 Gy; p<0.001), which resulted in a lower rate of ORN among those treated with IMPT (2.0% with IMPT vs.7.7% with IMRT).
Rwigema, et al. reported on the results of IMPT in 30 patients with oropharyngeal cancer treated with primary surgery followed by adjuvant proton therapy either or not combined with concurrent chemotherapy. The primary endpoints in this study were dysphagia, xerostomia, salivary duct inflammation and tube feeding dependence as assessed at 6 months after the end of treatment [Rwigema 2019]. In this study, a model-based clinical evaluation methodology was used [Langendijk 2018; Langendijk 2018], in which the observed toxicity rates after IMPT was compared to the predicted rates derived for the back-up IMRT-plans from the same 30 patients treated with IMPT, using validated Normal Tissue Complication Probability (NTCP) models. The NTCP-values for late dysphagia grade ≥ 2, dysphagia grade ≥ 3, tube feeding dependence, xerostomia grade ≥ 2 and salivary duct inflammation ≥ 2 based on the IMPT-plans were all significantly lower than based on the IMRT-plans, while the observed toxicity rates after IMPT were all equal or lower than predicted based on the NTCP-models.

In summary, limited data exists on the on late toxicity after protons as compared to photons. First results suggest less toxicity related to salivary function (xerostomia) and swallowing function, resulting from lower radiation exposure to relevant organs at risk, especially after longer follow up.

*Efficacy*
In total, seven studies were identified that compared efficacy endpoints between protons and photons. [Chou 2021; Li 2021; Yoon 2021; Alterio 2020; Park 2019; Blanchard 2016; Romesser 2016].

**Table 5: Summary of studies reporting on efficacy endpoints comparing photons with protons.**

| Reference | Tumour site | adiotherapy compariso | Numbers | Modalities | Endpoint | | |
|---|---|---|---|---|---|---|---|
| | | | | | LRC | PFS | OS |
| **Alterio 2020** | Nasopharynx | IMRT versus mixed beam | PhT: 17 patients PrT: 27 patients | Definitive RT with CRT and only 1 with RT alone | **2-year LRC**: PhT only: 92.4% PhT + PrT: 100% p=0.325 (logrank) | | |
| **Park 2019** | Nasopharynx | HT versus mixed beam | PhT: 63 patients PrT/PhT: 35 patients 2 x 35 patients used for | Concurrent CRT | | **1-year PFS**: PhT only: 82.9% PhT + PrT: 87.1% p=0.40 (logrank) | |
| **Chou 2021** | Nasopharynx | IMRT versus IMPT | PhT: 80 patients PrT: 80 patients | Concurrent CRT or induction CHT + RT or RT alone | | **2-year PFS**: PhT: 83.7% PrT: 94.4% p=0.10 (logrank) HR: 0.30 (CI: 0.09 - | **2-year OS**: PhT: 89.5% PrT: 100% p=0.10 (logrank) |
| **Li 2021** | Nasopharynx | IMRT versus IMPT | PhT: 49 patients PrT: 28 patients 2 x 24 patients used for PSM | Concurrent RT or RT alone | **2-year LRC**: PhT only: 86.2% PhT + PrT: 100% p=0.08 (logrank) | **2-year PFS**: PhT: 76.7% PrT: 95.7% HR: 0.33 (CI: 0.07 - | **3-year OS**: PhT: 94.1% PrT: 100% p=0.42 (logrank) |
| **Yoon 2021** | Oropharynx | IMRT versus mixed beam | PhT: 81 patients PrT: 67 patients 2 x 36 patients used for PSM | Concurrent CRT | | **2-year PFS**: PhT only: 78.8% PhT + PrT: 82.4% p=0.681 (logrank) | **2-year OS**: PhT only: 92.4% PhT + PrT: 100% p=0.325 (logrank) |
| **Blanchard 2016** | Oropharynx | IMRT versus IMPT | PhT: 100 patients PrT: 50 patients | Concurrent CRT or induction CHT + RT | **3-year LRC**: PhT: 89.7% PrT: 91.0% HR: 1.03 (CI: 0.35 - 3.02) | **3-year PFS**: PhT: 85.8% PrT: 86.4% HR: 1.00 (CI: 0.39 - 2.60) | **3-year OS**: PhT: 89.3% PrT: 94.3% HR: 0.55 (CI: 0.12 - 2.50) |
| **Romesser 2016** | Salivary gland and skin cancer | Unilateral IMRt versus uniform scanning protons | PhT: 23 patients PrT: 18 patients | Postoperative RT or CRT | **1-year LRC**: PhT: 95.5% PrT: 80.0% p=0.46 (logrank) | | **1-year OS**: PhT: 93.3% PrT: 83.3% p=0.08 (logrank) |

BioRT = bioradiation = radiotherapy + cetuximab
CRT = Chemoradiation
IMRT-SIB = Intensity Modulated Radiotherapy with Simultaneous Integrated Boost
LRC = Locoregional Control
PhT = Photon therapy
OS = Overall survival
PFS = Progression-free survival
PRO = Patient-rated Outcome
PrT = Proton therapy
PSM = Propensity score analysis
RT = Radiotherapy

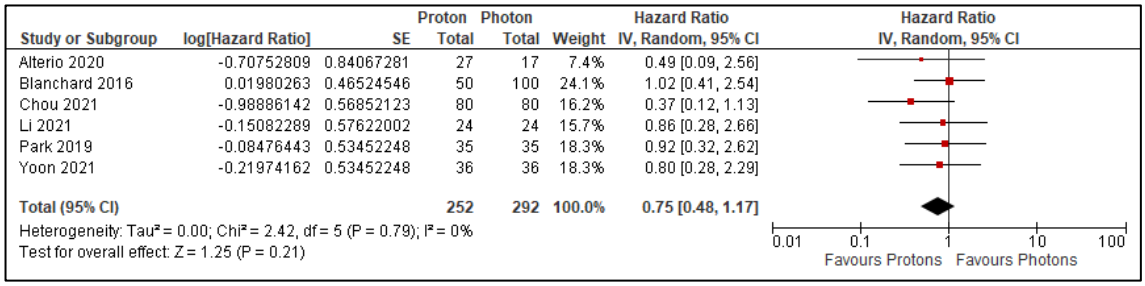
The characteristics of the selected studies are summarised in **Table 5**. Except for two studies [Romesser 2016, Alterio 2020], in most studies some kind of matching techniques were used to account for confounding, either based on propensity score matching [Park 2019; Yoon 2021; Li 2021; Chou 2021], or on predefined risk factors (Matched case control)[Blanchard 2016]. Some studies compared full treatment courses of proton therapy to photon therapy [Romesser 2016; Blanchard 2016; Li 2021; Chou 2021, while in others proton therapy group was used in combination photons (i.e., mixed beam technique) [Park 2019; Alterio 20-16; Yoon 2021].

Four of the selected studies described outcomes of nasopharyngeal cancer patients [Chou 2021; Li 2021; Park 2019; Alterio 2020], while in two studies, the population consisted of oropharyngeal cancer patients [Yoon 2021; Blanchard

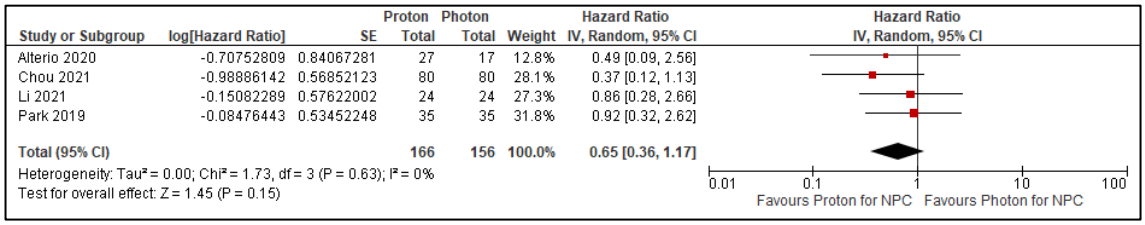
2016]. The remaining study included several types of head and neck cancer, mainly originating from the parotid or submandibular glands [Romesser 2016].

**Figure 10: Pooled analysis of progression-free survival, for all studies (A), studies including nasopharyngeal cancer patients (B), and studies including oropharyngeal cancer patients (C)**

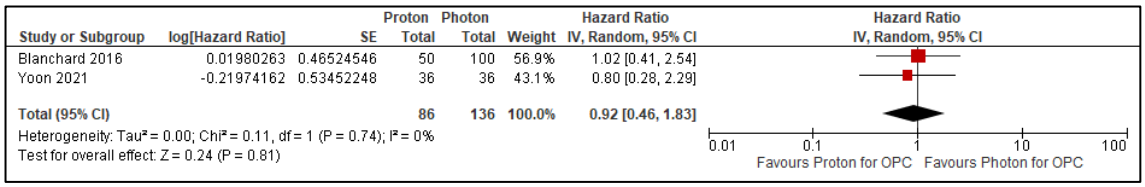
**A**




**B**




**C**




In none of the studies, significant differences were found regarding locoregional control, distant-metastases-free interval and/or overall survival between patients treated with protons versus those treated with photons. However, the majority of studies showed a trend in favour of proton therapy [Blanchard 2016; Park 2019; Alterio 2020; Yoon 2021; Li 2021; Chou 2021]. Progression-free survival reported for proton versus photon radiotherapy at 1, 2 and 3 years was 80.0-87.1% versus 82.9-95.5%,(4,6) 82.4-100% versus 76.7-86.2%,(7–10) and 86.4 versus 85.8%,(5) respectively.

Results of the pooled analysis on progression-free survival are summarized in **Figure 10**. The analysis showed a similar trend for progression-free survival, mainly for the studies including nasopharyngeal cancer patients, which was however not statistically significant. The HR for progression-free survival after proton versus photon radiotherapy for all nasopharyngeal and oropharyngeal cancer studies combined was 0.75 (95% CI 0.48-1.17; p=0.21). When performing separate meta-analyses for studies including nasopharyngeal and oropharyngeal

cancer patients, HRs were 0.65 (95% CI 0.36-1.17; p=0.15) and 0.92 (95% CI 0.46-1.83; 0.81), respectively.

## Discussion

The objective of this review was to investigate if the current literature supports the assumption that with proton therapy, toxicity can be significantly reduced without jeopardizing treatment efficacy.

In this review of literature, a significant reduction was observed regarding a number of acute toxicities. Many acute toxicities are primarily associated with acute mucositis, the most important dose limiting acute toxicity occurring during radiation in the head and neck region, which results in a number of subsequent other physician-rated toxicities and patient-rated complaints. In this review of current literature, a significant reduction of acute mucositis grade $\geq$ 3 was observed among patients treated with protons. Recently, we showed that the risk of acute mucositis depends on the mean dose to the oral cavity and treatment modality [van den Bosch 2021]. In most studies, radiotherapy with protons indeed resulted in a marked and statistically significant reduction of the dose to the oral cavity [Yasuda 2021; Romesser 2016; Li 2021; Holliday 2015; Cao 2021, Chiu 2021], which explains the lower rates of acute mucositis during proton therapy. Subsequently, the lower rates of acute mucositis may also at least partly explain lower rates of other toxicities observed after proton therapy, like dysphagia and weight loss. Moreover, a number of studies included in this review reported lower dose exposure with to the so-called swallowing organs-at-risk, like the pharyngeal constrictor muscles and the supraglottic area [Yasuda 2021; Alterio 2020, Chiu 2021], which may further contribute to the lower rates of dysphagia and patient-rated symptoms related to swallowing dysfunction, such as problems with swallowing, use of nutritional supplements and tube feeding dependence.

Limited data exists on the effect of late swallowing disorders (> 3 months after RT) of proton therapy. Given that the severity of acute dysphagia is very predictive for late dysphagia, it is likely that proton therapy eventually results in lower rates of late dysphagia as well [van der Laan 2015]. This is supported by the study of Rwigma, et al and Blanchard, et al, who found lower rates of dysphagia and tube feeding dependence at 6 months after treatment, and lower rates of grade 3 weight loss or tube feeding dependence at 6 and 12 months respectively [Rwigma 2019; Blanchard 2016].

Historically, new radiation technologies in head and neck cancer radiotherapy mainly focussed on preventing late xerostomia by reducing the dose to the salivary glands. A number of RCT's showed that reducing the dose to the parotid glands significantly reduced the rate of severe xerostomia in the first 24 months after RT [Nutting 2011]. With modern advanced photon techniques, like IMRT and VMAT, the dose to the parotid glands is already markedly reduced compared to the dose levels used in these RCT's. A striking finding was that IMPT seems to further reduce the risk of late xerostomia beyond 24 months, which was mainly related to a reduction of the dose to the oral cavity. With IMPT, the dose to the oral cavity can be reduced significantly, and in most studies, the dose difference between protons and photons is most pronounced to this organ-at-risk.

The results reported by Cao, et al, also indicate that longer follow up is needed to explore the full potential of proton therapy toxicity beyond 2 years after treatment.

The pooled analysis indicates higher rates of grade ≥ 3 acute dermatitis after proton therapy. This is not surprising as due to the beam properties of protons, the entry dose in the skin is generally higher than that of IMRT or VMAT. Normally, skin reactions after head and neck radiotherapy rapidly recover after the end of treatment, and in a recent analysis of our own institution, acute skin reactions 6 weeks after radiotherapy almost completely resolved and were not different anymore between photons and protons (unpublished data).

Recently published series showed an excellent progression-free survival after radiotherapy for head and neck cancer, with no significant differences between proton and photon therapy [Romesser 2-16; Blanchard 2016; Park 2019; Alterio 2020, Yoon 2021; Li 2021, Chou 2021] In most studies, a trend was observed towards better locoregional control after proton therapy [Blanchard 2016; Park 2019; Alterio 2020, Yoon 2021; Li 2021, Chou 2021]. The pooled analysis performed for this report showed a similar trend, in particular for patients with nasopharyngeal carcinoma, although it failed to reach statistical significance.
Some data exists on the radiobiological background of possible differences in radiobiological efficacy (RBE) between proton and photon therapy for head and neck cancer [Wang 2021]. Some studies found that cells irradiated with protons showed a gene/protein expression profile promoting cell kill and a more favourable inflammatory phenotype compared to those irradiated with photons, as well as reduced expression of genes involved in proliferation and (lymph-)angiogenesis [Wang 2021; Lupu-Plesu 2021; Wang 2020].Wang et al. found that head and neck squamous cell carcinoma cells had higher rates of unrepaired double-strand DNA breaks after irradiation with protons compared to photons [Wang 2017; Wang 2021]. The main types of cell death induced by irradiation of head and neck squamous cell cancer cells were mitotic catastrophe and cellular senescence, with a higher rate of both types of cell death after irradiation with protons compared to photons [Wang 2019; Wang 2021]. These results again highlight that protons interact with tumours and normal tissue in a different way, which may result in a different (potentially beneficial) therapeutic window.

The current analysis has a number of limitations.
First, results of RCT's are still lacking. To assess the possible beneficial effect of proton therapy on efficacy endpoints like locoregional control, progression-free survival and overall survival properly designed RCT's are required. At present, an RCT is ongoing in the United States, in which patients with HPV-positive oropharyngeal cancer treated with concurrent chemoradiation are randomly assigned to receive chemoradiation with either photons or protons. The primary endpoint in this non-inferiority study is progression-free survival. First results of this study are expected to become available in 2023. Although the results of RCT's are still lacking, the results of the current review of literature does not indicate that the efficacy is jeopardised as compared to photons.

It should be noted that there is less consensus how to assess the potential benefit of protons versus photons with regard to prevention of radiation-induced side effects [Langendijk 2013; Langendijk 2018]. For answering this question, RCT's are less suitable for a number of reasons. First, prevention of radiation-induced side effects mainly depends of the ability of a technology to reduce the dose to the organs-at-risk without jeopardising the target dose. RCT normally compare an experimental treatment with the current standard. The problem is that there is no current standard for any radiation technology. The eventual dose distributions on organs-at-risk do not only depend on the radiation technology itself, but particularly on how this technology is used [KNAW]: the so-called technology-user interplay. Second, radiation technologies continuously develop over time and have a typical life cycle of 5 to 7 years. This is nicely illustrated by the developments in proton therapy, which developed in one less than one decade from passive scattering to pencil beam scattering and is now at the phase to progress further into dynamic arc proton therapy, which further reduces the dose to multiple organs-at-risk at the same time [Liu 2020; Liu 2020].Consequently, given that most RCT's require 5 to 10 years from the initiation of the study until the publication of the final results, it is very unlikely that these studies will be practice changing as the results will be based on technologies that are considered outdated. In the Netherlands, an alternative methodology has been developed, referred to as the model-based approach, which consists of three main components, including model-based optimization, model-based selection (for protons) and model-based clinical evaluation [Langendijk 2018]. The model-based approach requires high quality NTCP-models, which are subject to the second part of this review.

Second, most studies reviewed in this report have relatively small sample sizes and limited follow-up duration. Consequently, conclusions regarding the possible differences of protons versus photons late toxicity and long-term efficacy cannot be determined.

Third, in three studies included in this analysis, mixed beam approaches were used, meaning that only a part of the treatment course was administered with protons and the largest part with photons. It is clear that such an approach will not explore the full potential of protons to widen the therapeutic window.

Fourth, there is major heterogeneity regarding the way outcome of protons versus photons is reported, which severely hampers data pooling, which could be a useful tool to get more insight in the potential benefits of protons versus photons. It is highly recommended to define guidelines for how to report outcome in these kind of studies, not only for RCT's but also for prospective observational cohort studies.

## Conclusions

At present, proton therapy in the treatment of HNC is mainly used to reduce the dose to the normal tissues with an equivalent dose to the target, aiming at reduction of radiation-induced side effects without jeopardising treatment efficacy. The results of the current literature review support that the dose reductions to organs-at-risk obtained with protons result in less acute radiation-induced toxicities and patient-reported outcome, without jeopardising locoregional control, progression-free survival and overall survival. To compare the efficacy of protons versus photons, RCT's are needed, while for assessing the benefit of protons to prevent radiation-induced side effects, alternative methodological approaches are required.

## References


Alterio D, D'Ippolito E, Vischioni B, Fossati P, Gandini S, Bonora M, et al. Mixed-beam approach in locally advanced nasopharyngeal carcinoma: IMRT followed by proton therapy boost versus IMRT-only. Evaluation of toxicity and efficacy. Acta oncologica (Stockholm, Sweden). 2020 May;59(5):541–8.

Blanchard P, Garden AS, Gunn GB, Rosenthal DI, Morrison WH, Hernandez M, et al. Intensity-modulated proton beam therapy (IMPT) versus intensity-modulated photon therapy (IMRT) for patients with oropharynx cancer - A case matched analysis. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2016 Jul;120(1):48–55.

Børvik E, Fjæra LF, Dahle TJ, Dale JE, Engeseth GM, Stokkevåg CH, et al. Exploration and application of phenomenological RBE models for proton therapy. Phys Med Biol 2018;63:185013 .

Chou Y-C, Fan K-H, Lin C-Y, Hung T-M, Huang B-S, Chang K-P, et al. Intensity Modulated Proton Beam Therapy versus Volumetric Modulated Arc Therapy for Patients with Nasopharyngeal Cancer: A Propensity Score-Matched Study. Cancers. 2021 Jul;13(14).

Ghosh-Laskar S, Yathiraj PH, Dutta D, Rangarajan V, Purandare N, Gupta T, Budrukkar A, Murthy V, Kannan S, Agarwal JP. Prospective randomized controlled trial to compare 3-dimensional conformal radiotherapy to intensity-modulated radiotherapy in head and neck squamous cell carcinoma: Long-term results. Head Neck. 2016 Apr;38 Suppl 1:E1481-7.

ICRU. Prescribing, Recording, and Reporting Proton-Beam. ICRU Rep 78. vol. 7. 2007. doi:10.1093/jicru/ndi004.

Jakobi A, Stützer K, Bandurska-Luque A, Löck S, Haase R, Wack LJ, Mönnich D, Thorwarth D, Perez D, Lühr A, Zips D, Krause M, Baumann M, Perrin R, Richter C. NTCP reduction for advanced head and neck cancer patients using proton therapy for complete or sequential boost treatment versus photon therapy. Acta Oncol. 2015;54(9):1658-64.

Kam MK, Leung SF, Zee B, Chau RM, Suen JJ, Mo F, Lai M, Ho R, Cheung KY, Yu BK, Chiu SK, Choi PH, Teo PM, Kwan WH, Chan AT. Prospective randomized study of intensity-modulated radiotherapy on salivary gland function in early-stage nasopharyngeal carcinoma patients. J Clin Oncol. 2007 Nov 1;25(31):4873-9.

KNAW (2014). Royal Academy of Arts and Sciences. Evaluation of new technology in Health care. In need of guidance for relevant evidence. Amsterdam (KNAW). ISBN 978-90-6984-686-6.

Langendijk JA, Boersma LJ, Rasch CRN, van Vulpen M, Reitsma JB, van der Schaaf A, Schuit E. Clinical Trial Strategies to Compare Protons With Photons. Semin Radiat Oncol. 2018 Apr;28(2):79-87.

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 Aug 1;26(22):3770-6.

Langendijk JA, Lambin P, De Ruysscher 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 Jun;107(3):267-73.

Langendijk JA, Orecchia R, Haustermans K, Zips D, Balosso J, Lacombe D, Lievens Y, Weber DC, Grau C, Troost EGC. Prospective data registration and clinical trials for particle therapy in Europe. Radiother Oncol. 2018 Jul;128(1):9-13.

Li X, Kitpanit S, Lee A, Mah D, Sine K, Sherman EJ, et al. Toxicity Profiles and Survival Outcomes Among Patients With Nonmetastatic Nasopharyngeal Carcinoma Treated With Intensity-Modulated Proton Therapy vs Intensity-Modulated Radiation Therapy. JAMA network open. 2021 Jun;4(6):e2113205.

Liu G, Li X, Qin A, Zheng W, Yan D, Zhang S, Stevens C, Kabolizadeh P, Ding X. Improve the dosimetric outcome in bilateral head and neck cancer (HNC) treatment using spot-scanning proton arc (SPArc) therapy: a feasibility study. Radiat Oncol. 2020 Jan 30;15(1):21.

Liu G, Li X, Zhao L, Zheng W, Qin A, Zhang S, Stevens C, Yan D, Kabolizadeh P, Ding X. A novel energy sequence optimization algorithm for efficient spot-scanning proton arc (SPArc) treatment delivery. Acta Oncol. 2020 Oct;59(10):1178-1185.

Lupu-Plesu M, Claren A, Martial S, N'Diaye P-D, Lebrigand K, Pons N, et al. Effects of proton versus photon irradiation on (lymph)angiogenic, inflammatory, proliferative and anti-tumor immune responses in head and neck squamous cell carcinoma. Oncogenesis. 2017 Jul;6(7):e354.

Matthiesen C, Herman Tde L, Singh H, Mascia A, Confer M, Simpson H, Higby C, Arain A, Keole S, Herman T, Bogardus C, Zhao YD, Ahmad S.

Dosimetric and radiobiologic comparison of 3D conformal, IMRT, VMAT and proton therapy for the treatment of early-stage glottic cancer. J Med Imaging Radiat Oncol. 2015 Apr;59(2):221-8.

Meijer TWH, Scandurra D, Langendijk JA. Reduced radiation-induced toxicity by using proton therapy for the treatment of oropharyngeal cancer. The British journal of radiology. 2020 Mar;93(1107):20190955.

Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, Miles EA, Miah AB, Newbold K, Tanay M, Adab F, Jefferies SJ, Scrase C, Yap BK, A'Hern RP, Sydenham MA, Emson M, Hall E; PARSPORT trial management group. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. Lancet Oncol. 2011 Feb;12(2):127-36.

Paganetti H, Niemierko A, Ancukiewicz M, Gerweck LE, Goitein M, Loeffler JS, et al. Relative biological effectiveness (RBE) values for proton beam therapy. Int J Radiat Oncol 2002;53:407–21. doi:10.1016/S0360-3016(02)02754-2.

Paganetti H. Relative biological effectiveness (RBE) values for proton beam therapy. Variations as a function of biological endpoint, dose, and linear energy transfer. Phys Med Biol 2014;59

Park SG, Ahn YC, Oh D, Noh JM, Ju SG, Kwon D, et al. Early clinical outcomes of helical tomotherapy/intensity-modulated proton therapy combination in nasopharynx cancer. Cancer science. 2019 Sep;110(9):2867–74.

Romesser PB, Cahlon O, Scher E, Zhou Y, Berry SL, Rybkin A, et al. Proton beam radiation therapy results in significantly reduced toxicity compared with intensity-modulated radiation therapy for head and neck tumors that require ipsilateral radiation. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2016 Feb;118(2):286–92.

Rwigema JM, Langendijk JA, Paul van der Laan H, Lukens JN, Swisher-McClure SD, Lin A. A Model-Based Approach to Predict Short-Term Toxicity Benefits With Proton Therapy for Oropharyngeal Cancer. Int J Radiat Oncol Biol Phys. 2019 Jul 1;104(3):553-562.

Swisher-McClure S, Teo BK, Kirk M, Chang C, Lin A. Comparison of Pencil Beam Scanning Proton- and Photon-Based Techniques for Carcinoma of the Parotid. Int J Part Ther. 2016 Mar;2(4):525-532.

Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials. 2007 Jun;8:16.

Toledano I, Graff P, Serre A, Boisselier P, Bensadoun RJ, Ortholan C, Pommier P, Racadot S, Calais G, Alfonsi M, Favrel V, Giraud P, Lapeyre M. Intensity-modulated radiotherapy in head and neck cancer: results of the

prospective study GORTEC 2004-03. Radiother Oncol. 2012 Apr;103(1):57-62.

Van den Bosch L, van der Schaaf A, van der Laan HP, Hoebers FJP, Wijers OB, van den Hoek JGM, Moons KGM, Reitsma JB, Steenbakkers RJHM, Schuit E, Langendijk JA. Comprehensive toxicity risk profiling in radiation therapy for head and neck cancer: A new concept for individually optimised treatment. Radiother Oncol. 2021 Apr;157:147-154.

van der Laan HP, Bijl HP, Steenbakkers RJ, van der Schaaf A, Chouvalova O, Vemer-van den Hoek JG, Gawryszuk A, van der Laan BF, Oosting SF, Roodenburg JL, Wopken K, Langendijk JA. Acute symptoms during the course of head and neck radiotherapy or chemoradiation are strong predictors of late dysphagia. Radiother Oncol. 2015 Apr;115(1):56-62.

Wang L, Fossati P, Paganetti H, Ma L, Gillison M, Myers JN, et al. The Biological Basis for Enhanced Effects of Proton Radiation Therapy Relative to Photon Radiation Therapy for Head and Neck Squamous Cell Carcinoma. International journal of particle therapy. 2021;8(1):3–13.

Wang L, Han S, Zhu J, Wang X, Li Y, Wang Z, et al. Proton versus photon radiation-induced cell death in head and neck cancer cells. Head & neck. 2019 Jan;41(1):46–55.

Wang L, Wang X, Li Y, Han S, Zhu J, Wang X, et al. Human papillomavirus status and the relative biological effectiveness of proton radiotherapy in head and neck cancer cells. Head & neck. 2017 Apr;39(4):708–15.

Wang L, Yang L, Han S, Zhu J, Li Y, Wang Z, et al. Patterns of protein expression in human head and neck cancer cell lines differ after proton vs photon radiotherapy. Head & neck. 2020 Feb;42(2):289–301.

Wedenberg M, Lind BK, Hårdemark B. A model for the relative biological effectiveness of protons: the tissue specific parameter α/β of photons is a predictor for the sensitivity to LET changes. Acta Oncol 2013;52:580–8.

Widesott L, Pierelli A, Fiorino C, Dell'oca I, Broggi S, Cattaneo GM, Di Muzio N, Fazio F, Calandrino R, Schwarz M. Intensity-modulated proton therapy versus helical tomotherapy in nasopharynx cancer: planning comparison and NTCP evaluation. Int J Radiat Oncol Biol Phys. 2008 Oct 1;72(2):589-96.

Yoon HG, Ahn YC, Oh D, Noh JM, Park SG, Nam H, et al. Early Clinical Outcomes of Intensity Modulated Radiation Therapy/Intensity Modulated Proton Therapy Combination in Comparison with Intensity Modulated Radiation Therapy Alone in Oropharynx Cancer Patients. Cancers. 2021 Mar;13(7).

Zhang W, Zhang X, Yang P, Blanchard P, Garden AS, Gunn B, Fuller CD, Chambers M, Hutcheson KA, Ye R, Lai SY, Radwan MAS, Zhu XR, Frank SJ. Intensity-modulated proton therapy and osteoradionecrosis in oropharyngeal cancer. Radiother Oncol. 2017 Jun;123(3):401-405.

# TASK 1.4.2: REVIEW ON PREDICTION MODELS THAT PREDICT RADIATION-INDUCED TOXICITIES IN PATIENTS WITH HEAD AND NECK CANCER

## EXECUTIVE SUMMARY

**Background** Multivariable prediction models have been introduced to predict a patient's normal tissue complication probability (NTCP) based on patient, disease, and treatment characteristics, including radiation dose to healthy tissues. These can assist clinicians when evaluating and choosing the optimal treatment plan among various conventional and emerging techniques, such as proton therapy, by comparing the predicted complication risk of each treatment plan. Various prognostic models for prediction of various types of radiation-induced side effects in patients with head and neck cancer have been developed but an overview of all available prognostic models predicting radiation-induced side effects, including an appraisal of their quality and applicability is currently lacking.

**Methods** We performed a systematic review to assess prognostic models available to predict the risk of radiation-induced side effects after radiation exposure to patients with head and neck cancer, and assess their predictive performance, quality, and applicability. We searched in Ovid MEDLINE and Ovid Embase to identify development or validation studies of any type of prognostic models that predicted acute or late radiation-induced side effects in patients with head and neck cancer undergoing current standard radiotherapy techniques (e.g., intensity modulated radiotherapy [IMRT] or volumetric modulated arc therapy [VMAT], 3D conformal radiotherapy and/or proton therapy). Title, abstract, and full text screening, and data extraction (based on the CHARMS checklist) was performed in duplicate. Risk of bias and applicability was assessed using PROBAST.

**Results** We identified 242 models that were developed and 64 external validation studies, mostly for the prediction of toxicities related to saliva and swallowing. The number of external validation studies ranged between 1 to 4 for models with an external validation. In most studies, inclusion criteria were poorly reported, and a proper reporting of the model performance was lacking. Ninety-seven percent of both the model development and external validation studies had an overall high risk of bias according to PROBAST, mainly due to a high risk of bias in the analysis domain, while applicability concerns were mostly low in model development studies (63%) but often unclear in external validation studies (59%). Only four models in the development studies had a low risk of bias (3 low and 1 with unclear applicability concern) whereas only 1 external validation study was assessed as having low risk of bias and applicability concern.

**Conclusion** Although many models have been developed and a minority has been validated, most of the models cannot be used to provide an approximate

radiotherapy-induced toxicity prediction but require testing and updating before being applied to new patients.

# PROGNOSTIC MODELS FOR RADIATION-INDUCED COMPLICATIONS AFTER RADIOTHERAPY IN HEAD AND NECK CANCER PATIENTS

## A SYSTEMATIC REVIEW

### Introduction

Radiotherapy is performed in over 60% of cancer patients(Halperin, 2008) and is a balance between destroying the tumour (tumour control) and preserving the surrounding normal tissues (toxicity minimization). As these toxicities have a significant impact on the more general dimensions of quality of life(Langendijk et al., 2008), prevention of these side effects is crucial. Historically, toxicity has been minimized by setting maximum radiation dose thresholds to organs (e.g., 54 Gray to the brainstem).(Emami, 2013) More recently, more multivariable approaches have been introduced to predict a patient's normal tissue complication probability (NTCP) based on patient, disease, and treatment characteristics, including radiation dose to healthy tissues. These NTCP models can be used during treatment plan optimization to actively guide the dose distribution to lower the complication risk.(Kierkels et al., 2014) Additionally, they can assist clinicians when evaluating and choosing the optimal treatment plan among various conventional and emerging techniques, such as proton therapy, by comparing the predicted complication risk of each treatment plan.(Langendijk et al., 2013; Widder et al., 2016) In the latter situation, a certain threshold for the difference in NTCPs between both therapies could be defined, i.e., a ΔNTCP, above which a new treatment could be indicated. Since both therapies can, theoretically, provide the same tumour control, the ΔNTCP is caused by the difference in radiation dose to normal tissue between the two therapies. Various prognostic models for prediction of various types of radiation-induced side effects in patients with head and neck cancer (e.g. xerostomia, dysphagia, hearing loss, and hypothyroidism), have been developed (Beetz et al., 2012; Cheraghi et al., 2017; Christianen et al., 2016; Luo et al., 2018; Wopken et al., 2014), but an overview of all available prognostic models predicting radiation-induced side effects, including an appraisal of their quality and applicability is currently lacking.

*Objective*

To assess prognostic models available to predict the risk of radiation-induced side effects after radiation exposure to patients with head and neck cancer, and assess their predictive performance, quality, and applicability.

## Implementation

A detailed description of the methods can be found in the Cochrane protocol of this review.(Takada et al., 2021)

### Criteria for considering studies for this review

Types of studies

Based on the PICOTS (Population, Intervention, Comparison, Outcome, Timing) of the review question (Table 1)(Debray et al., 2017; K. G. Moons et al., 2014; K. G. M. Moons et al., 2019; Wolff et al., 2019), we included studies which met the following criteria:

iv) study design: all retrospective and prospective cohort and nested case-control studies,

v) data source: studies that used routine care, registry data or data from randomized trials,

vi) aim: studies that aimed to develop, evaluate (internal and/or external validation) or update prognostic models to predict radiation-induced side effects in patient with head and neck cancer who underwent radiotherapy. The timing of model usage is just before starting radiotherapy.

There was no restriction on language.

**Table 1 PICOTS of the review question based on the CHARMS checklist**

| | |
|---|---|
| Population targeted | Patients with head and neck cancer undergoing current standard radiotherapy techniques (e.g., intensity modulated radiotherapy [IMRT] or volumetric modulated arc therapy [VMAT], 3D conformal radiotherapy and/or proton therapy) |
| Index model(s) | All available prognostic models predicting the risk of radiation-induced side effects |
| Comparator model(s) | Not applicable |
| Outcome(s) to be predicted | All types of acute and late radiation-induced side effects in head and neck regions |
| Timing of making prediction and Time span of the prediction | Just before starting radiotherapy<br><br>There were no restrictions on the prediction horizon |
| Setting | Secondary and tertiary care |

### Search methods for identification of studies

Electronic searches

We conducted the search in the following databases: Ovid MEDLINE and Ovid Embase. The search strategies for each database have been presented in the review protocol [ref Takada Cochrane protocol]. To efficiently identify prognostic model studies, we used and modified the search filter described by Geersing and colleagues(Geersing et al., 2012) for our purpose.

Searching other resources

We searched the following databases for ongoing trials: ClinicalTrials.gov and World Health Organization International Clinical Trials Registry Platform (WHO ICTRP). We checked reference articles cited in the retrieved articles. In addition, clinical experts in the field were contacted to identify any relevant prognostic models we may have missed with our search. Models published in grey literature (i.e., conference abstracts) were not considered for inclusion.

**Data collection**

Selection of studies

Six review authors independently screened the search results for eligibility on title and abstract. Then, the same review authors independently assessed the eligibility of potentially relevant studies by reading the full-text articles. Any disagreement between paired review authors was resolved by discussion and if needed by consulting an independent review author (Ewoud Schuit). We documented study selection in a flow chart as recommended in the PRISMA guidelines.(Moher, Liberati, Tetzlaff, Altman, & Group, 2009)

Data extraction and management

The same review authors independently extracted the data in accordance with the CHARMS checklist.(K. G. Moons et al., 2014) A detailed list of items is provided in the review protocol.(Takada et al., 2021) These items provide information to assess applicability to the review question and their risks of bias (see below). Any disagreement between the review authors was resolved by discussion and if needed by consulting an independent review author (ES). We contacted authors of individual studies to obtain additional information, if necessary.

Assessment of risk of bias and applicability

We used the PROBAST-tool for assessment of risk of bias and applicability of all models reported in the included studies.(K. G. M. Moons et al., 2019; Wolff et al., 2019) Risk of bias and applicability was assessed in the participants, predictors, outcome, and analysis domain as high, low or unclear risk of bias. Concerns regarding applicability were rated similarly to risk of bias, but without signalling questions. The same review authors independently assessed risk of bias and applicability. Any disagreement between the authors was resolved by discussion or by consulting a third review author (ES).

**Data synthesis**

Data synthesis and meta-analysis approaches

Data analysis was performed using descriptive statistics.

# Results

## 1. Results of the search

A total of 10119 unique records were identified by our search after removal of the duplicates. An additional 55 records were identified by tracing clinicaltrial.gov and World Health Organization International Clinical Trials Registry Platform (WHO ICTRP). Of 10174 records, 9421 were excluded based on title and abstract screening, leaving 753 records for full text screening. Based on the full text screening, 655 publications were excluded which reported a conference abstract (n = 214), not an original study (n = 4), no prediction model or a predictor finding study (n = 325), wrong domain (n = 62), wrong outcome (n = 18) and others (n = 32) (Figure 1). From the remaining 98 publication included in the data extraction. In total data was extracted for 242 developed models and external validation of 64 models from these 98 publications (172 models from 51 articles reporting a model development study, 63 models from 27 articles reporting a model development and internal validation study, 5 models from 2 articles reporting a model development and external validation study, 2 models from 2 articles reporting a model development and both internal and external validation study and finally 57 models from 16 articles reporting an external validation study) (Figure 1).

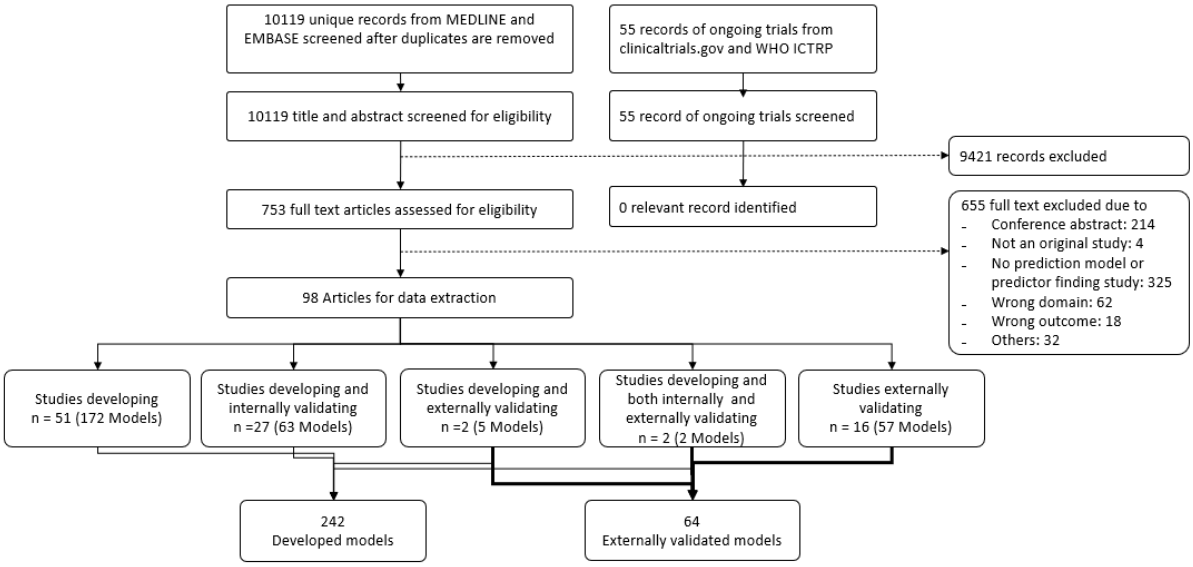



**Figure 1. PRISMA study workflow**.

The results for model development and external validation studies are presented separately.

## 2. Model development studies

Of the 242 models included, 172 (71%) were reported in publications with a model development study only, 63 (26%) models were reported in model development with an internal validation study, 7 (3%) were reported in model development and external validation study.

Most of the models were developed in prospective cohort studies (n= 143 models, 59%) involving generally <5 centres (n =199 models, 88%) from Europe (n = 86 models, 36%) followed by North America (n = 72, 30%) and Asia (n= 66, 28%).

Patient recruitment method was consecutive in 88 models (36%) and unclear in 151 models (62%).

### 2.1. Study inclusion criteria

The inclusion criteria of the 242 models and three most modelled outcomes (swallowing-related, salivary-related, and brain and nerve-related) are given in the Table 2.

Of the 242 models included, tumour site was specified in the study inclusion criteria as multiple tumour sites for 53 models (22%), followed by nasopharynx (n= 32, 13%), oropharynx (n=30, 12%), and not specified for 122 models (50%). Similarly, histological tumour type (n= 175, 72%) and tumour stage (n= 186, 77%) were not specified in study inclusion criteria in majority of the models. Consequently, most models were not targeted to specific tumour sites, histological tumour type or stage.

The most common RT techniques were IMRT (including VMAT and tomotherapy) (n = 84, 35%) while in 80 models (33%) RT technique was not specified in study inclusion criteria. In addition, neither RT target dose (n = 100, 41%) nor RT fractionation scheme (n=136, 56%) were specified. Furthermore, other treatment modalities including surgery (n= 193, 80%), chemotherapy (n=148, 61%) and molecular therapy such as cetuximab (n=215, 89%) were not mentioned in the study inclusion criteria for majority of the models, meaning most models focused on a broader head and neck patient population.

**Table 2. Characteristics of study inclusion criteria for all models and group of toxicities with the highest number of models in model development studies with or without an internal validation.**

| Variable n (%) | All models | Models for toxicities related to saliva | Models for toxicities related to swallowing | Models for toxicities related to brain and nerve injury |
|---|---|---|---|---|
| **N of models** | 242 (100) | 109 (100) | 44 (100) | 20 (100) |
| **Aim of the publication** | | | | |
| Development only | 172 (71.1) | 81 (74.3) | 33 (75.0) | 16 (80.0) |
| Development + internal val. | 63 (26.0) | 22(20.2) | 10(22.7) | 4 (20.0) |
| Development + external val. | 5 (2.1) | 4(3.7) | 1 (2.3) | 0 (0.0) |
| Development + internal + external val. | 2 (0.8) | 2 (1.8) | 0 (0.0) | 0 (0.0) |
| **Region** | | | | |
| Europe | 86 (36.0) | 32 (29.6) | 18 (41.9) | 1 (5.0) |
| North America | 72 (30.1) | 42 (38.9) | 18 (41.9) | 1 (5.0) |
| Asia | 66 (27.6) | 32 (29.6) | 0 (0.0) | 18 (90.0) |
| Combination | 13 (5.4) | 2 (1.9) | 5 (11.6) | 0 (0.0) |
| Australia | 2 (0.8) | 0 (0) | 2 (4.7) | 0 (0.0) |
| Not reported | 3 (1.4) | 1 (0.9) | 1 (2.3) | 0 (0.0) |
| **N of centres** | | | | |
| < 5 | 199 (88.4) | 88 (88.0) | 36 (81.8) | 17 (85.0) |
| 5-9 | 14 (6.2) | 11 (11.0) | 0 (0.0) | 1 (5.6) |
| > 10 | 12 (5.3) | 1 (1.0) | 5 (12.2) | 0 (0.0) |

| | | | | |
|---|---|---|---|---|
| Not reported | 17 (7.0) | 9 (8.3) | | 2 (10.0) |
| **Design** | | | | |
| Prospective cohort | 143 (59.1) | 86 (78.9) | 26 (59.1) | 7 (35.0) |
| Retrospective cohort | 62 (25.6) | 9 (8.3) | 7 (15.9) | 13 (65.0) |
| Randomised trial participants | 33 (13.6) | 13 (11.9) | 10 (22.7) | 0 (0.0) |
| Unclear | 4 (1.7) | 1 (0.9) | 1 (2.3) | 0 (0.0) |
| **Patient recruitment** | | | | |
| Consecutive | 88 (36.4) | 19 (17.4) | 16 (36.4) | 10 (50.0) |
| Non-consecutive | 3 (1.2) | 2 (1.8) | 1 (2.3) | 0 (0.0) |
| Unclear | 151 (62.4) | 88 (80.7) | 27 (61.4) | 10 (50.0) |
| ***Study inclusion criteria*** | | | | |
| **Tumour Site** | | | | |
| Multiple sites | 53 (21.9) | 26 (23.9) | 11 (25.0) | 1 (5.0) |
| Nasopharynx | 32 (13.2) | 14 (12.8) | 0 (0.0) | 6 (30.0) |
| Oropharynx | 30 (12.4) | 12 (11.0) | 18 (40.9) | 0 (0.0) |
| Soft tissue | 3 (1.2) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Hypopharynx | 2 (0.8) | 0 (0.0) | 0 (0.0) | 2 (10.0) |
| Non-specified | 122 (50.4) | 57 (52.3) | 15 (34.1) | 11 (55.0) |
| **Cancer type** | | | | |
| Non-specified | 175 (72.3) | 79 (72.5) | 26 (59.1) | 13 (65.0) |
| SCC | 55 (22.7) | 30 (27.5) | 16 (36.4) | 1 (5.0) |
| Others | 12 (5.0) | 0 (0.0) | 2 (4.5) | 6 (30.0) |
| **Cancer stage** | | | | |
| Specified in inclusion criteria | 56 (23.1) | 22 (20.2) | 14 (31.8) | 7 (35.0) |
| Non-specified | 186 (76.9) | 87 (79.8) | 30 (68.2) | 13 (65.0) |
| **Type of radiotherapy technique** | | | | |
| IMRT (incl. VMAT and tomotherapy) | 84 (34.7) | 36 (33.0) | 20 (45.5) | 9 (45.0) |
| Combination of different techniques | 56 (23.1) | 41 (37.6) | 11 (25.0) | 0 (0.0) |
| 3D conformal | 20 (8.3) | 9 (8.3) | 0 (0.0) | 0 (0.0) |
| SBRT | 1 (0.4) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Particle therapy | 1 (0.4) | 0 (0.0) | 0 (0.0) | 1 (5.0) |
| Non-specified | 80 (33.1) | 23 (21.1) | 13 (29.5) | 10 (50.0) |
| **Target radiation dose** | | | | |
| ≥ 66 Gy | 81 (33.5) | 43 (39.4) | 15 (34.1) | 6 (30.0) |
| Case mix of different target doses | 35 (14.5) | 14 (12.8) | 3 (6.8) | 1 (5.0) |
| < 66 Gy | 26 (10.7) | 14 (12.8) | 9 (20.5) | 0 (0.0) |
| Non-specified | 100 (41.3) | 38 (34.9) | 17 (38.6) | 13 (65.0) |
| **Radiotherapy fractionation** | | | | |
| Conventional fractionation | 77 (31.8) | 48 (44.0) | 6 (13.6) | 8 (40.0) |
| Non-conventional fractionation | 28 (11.6) | 6 (5.5) | 11 (25.0) | 1 (5.0) |
| Case mix with different fractionations | 1 (0.4) | 0 (0.0) | 1 (2.3) | 0 (0.0) |
| Non-specified | 136 (56.2) | 55 (50.5) | 26 (59.1) | 11 (55.0) |
| **Type of surgery** | | | | |
| No surgery | 34 (14.0) | 10 (9.2) | 5 (11.4) | 0 (0.0) |
| Case mix | 13 (5.4) | 9 (8.3) | 3 (6.8) | 0 (0.0) |
| Post-surgery | 2 (0.8) | 2 (1.8) | 0 (0.0) | 0 (0.0) |
| Non-specified | 193 (79.8) | 88 (80.7) | 36 (81.8) | 20 (100.0) |
| **Chemotherapy** | | | | |
| Yes | 91 (37.6) | 28 (25.7) | 32 (72.7) | 8 (40.0) |

| | | | | |
|---|---|---|---|---|
| No chemotherapy | 3 (1.2) | 0 (0.0) | 1 (2.3) | 0 (0.0) |
| Non-specified | 148 (61.2) | 81 (74.3) | 11 (25.0) | 12 (60.0) |
| **Molecular therapy** | | | | |
| No molecular therapy | 18 (7.5) | 7 (6.4) | 3 (6.8) | 3 (15.0) |
| Yes | 9 (3.8) | 5 (4.6) | 1 (2.3) | 0 (0.0) |
| Non-specified | 215 (88.8) | 97 (89.0) | 40 (90.9) | 17 (85.0) |

## 2.2. Characteristics of models

Almost half of the models were developed to predict toxicities related to salivary (n= 109, 45%), followed by swallowing-related complications (n = 44, 18%), brain and nerve toxicities (n = 20, 8%) and hypothyroidism (n = 15, 6%). Logistic regression (n= 89, 37%) and Lyman Kutcher Burman (n =28, 12%) modelling was the most common modelling technique.

The age of the participants ranged between 10.2 – 67.0 years while for 142 models (59%), the age of the participants was not reported. The proportion of the male participants varied between 26.7% – 97% in 171 models (71%) in which the gender of the participants was reported.

The number of events per variable were ≥20 in 44 (18%), 10-19 in 27 (11%) and <10 and 63 (32%) models. The handling of the missing data was not reported in most studies (n = 138, 57%) while the most common method was complete case analysis (n = 82, 34%). The internal validation technique was rarely performed (n = 65 models, 27%) and done mostly using resampling the same data set (n = 37, 15%). Most models were presented as a full mathematical equation including both intercept/baseline hazard and coefficients (n = 153, 63%).

Calibration and c-statistics of the apparent model were reported only in 59 (24%) and 92 (38%) models. The c-statistics values ranged between 0.60 and 0.98 (0.66 – 0.98 in models for toxicities related to salivary, 0.67 – 0.88 in models for toxicities related to swallowing and 0.68 – 0.95 in models for toxicities related to brain and nerve injury). Furthermore, of the 65 models with an internal validation, calibration after internal validation was reported only in 15 (23%) models while c-statistics after internal validation were reported in 50 (77%) models, which varied between 0.49 and 0.90 (0.55 – 0.90 in 18 models for toxicities related to salivary, 0.71 – 0.86 in 9 models for toxicities related to swallowing and 0.68 – 0.78 in 4 models for toxicities related to brain and nerve injury).

The characteristics of the 242 models and three most modelled outcomes (swallowing-related, salivary-related, and brain and nerve-related) were given in the Table 3.

**Table 3. Characteristics of all models and group of toxicities with the highest number of models in model development studies with or without an internal validation.**

| Variable n (%) | All models | Models for toxicities related to saliva | Models for toxicities related to swallowing | Models for toxicities related to brain and nerve injury |
|---|---|---|---|---|
| **N of models** | 242 (100) | 109 (100) | 44 (100) | 20 (100) |

| | | | | |
|---|---|---|---|---|
| **Outcome** | | | | |
| Saliva-related | 109 (45.0) | 109 (100) | | |
| Swallowing-related | 44 (18.2) | | 44 (100) | |
| Brain and nerve-related | 20 (8.3) | | | 20 (100) |
| Hypothyroidism | 15 (6.2) | | | |
| Mucosa-related | 14 (5.8) | | | |
| Trismus | 13 (5.4) | | | |
| Lethal toxicities | 10 (4.1) | | | |
| Weight loss and denutrition | 7 (2.9) | | | |
| Neck fibrosis | 5 (2.1) | | | |
| Skin-related | 3 (1.2) | | | |
| Laryngeal toxicities | 1 (0.4) | | | |
| Any late toxicity | 1 (0.4) | | | |
| **Type of model** | | | | |
| Logistic regression | 89 (36.8) | 39 (35.8) | 22 (50.0) | 3 (15.0) |
| Lyman Kutcher Burman model | 28 (11.6) | 16 (14.7) | 2 (4.5) | 3 (15.0) |
| Machine learning | 16 (6.6) | 1 (0.9) | 4 (9.1) | 0 (0.0) |
| Time-to-event model | 8 (3.3) | 4 (3.7) | 0 (0.0) | 2 (10.0) |
| Others | 96 (39.7) | 47 (43.1) | 16 (36.4) | 9 (45.0) |
| Unclear | 5 (2.1) | 2 (1.8) | 0 (0.0) | 3 (15.0) |
| **Age** | | | | |
| Not reported | 142 (58.7) | 66 (60.6) | 28 (63.6) | 16 (80.0) |
| Range | 10.2 - 67.0 | 46.4 – 61.0 | 55.0 – 63.0 | 46.0 – 62.0 |
| **Proportion of male** | | | | |
| Not reported | 71 (29.3) | 29 (26.7) | 17 (38.6) | 11 (55.0) |
| Range | 26.7 - 97.0 | 66.0 – 97.0 | 58.5 – 90.7 | 53.0 – 90.3 |
| **Event per candidate variable** | | | | |
| ≥ 20 | 44 (18.2) | 16 (11.9) | 5 (11.4) | 10 (50.0) |
| ≥ 10 & < 20 | 27 (11.2) | 13 (14.7) | 6 (13.6) | 0 (0.0) |
| < 10 | 63 (31.7) | 30 (27.5) | 18 (47.4) | 5 (25.0) |
| Not applicable | 65 (32.7) | 20 (18.3) | 9 (23.7) | 4 (20.0) |
| Not reported | 43 (17.8) | 30 (27.5) | 6 (13.6) | 1 (5.0) |
| **Handling of missing values** | | | | |
| Complete case analysis | 82 (33.9) | 27 (24.8) | 23 (52.3) | 2 (10.0) |
| Single imputation | 1 (0.4) | 0 (0.0) | 1 (2.3) | 0 (0.0) |
| Multiple imputation | 1 (0.4) | 0 (0.0) | 0 (0.0) | 1 (5.0) |
| Others | 20 (8.3) | 15 (13.8) | 1 (2.3) | 0 (0.0) |
| Non-specified | 138 (57.0) | 67 (61.5) | 19 (43.2) | 17 (85.0) |
| **Technique for internal validation** | | | | |
| Resampling of same data set | 37 (15.3) | 17 (15.6) | 6 (13.6) | 0 (0.0) |
| Cross-validation | 12 (5.0) | 3 (2.8) | 4 (9.1) | 1 (5.0) |
| Random split of data set | 12 (5.0) | 3 (2.8) | 0 (0.0) | 2 (10.0) |
| Non-random split of data set | 2 (0.8) | 1 (0.9) | 0 (0.0) | 1 (5.0) |
| No internal validation performed | 177 (73.1) | 85 (78.0) | 34 (77.3) | 16 (80.0) |
| Unclear | 2 (0.8) | 0 (0.0) | 0 (0) | 0 (0.0) |
| **Type of model presentation** | | | | |
| Full mathematical equation | 153 (63.2) | 78 (71.6) | 22 (50.0) | 11 (55.0) |
| Part of mathematical equation | 27 (11.2) | 10 (9.2) | 5 (11.4) | 4 (20.0) |
| Graphical presentation | 19 (7.9) | 1 (0.9) | 9 (20.5) | 4 (20.0) |
| Scoring system | 8 (3.3) | 4 (3.7) | 3 (6.8) | 0 (0.0) |
| Non-specified | 35 (14.5) | 16 (14.7) | 5 (11.4) | 1 (5.0) |

| | | | | |
|---|---|---|---|---|
| **Apparent model performance** | | | | |
| Calibration reported | 59 (24.4) | 18 (16.5) | 7 (15.9) | 9 (45.0) |
| c-statics reported | 92 (38.0) | 36 (33.0) | 23 (52.3) | 9 (45.0) |
| c-statics (range) | 0.60 - 0.98 | 0.66 - 0.98 | 0.67 - 0.88 | 0.68 - 0.95 |
| **Model performance after internal validation*** | | | | |
| Calibration reported | 15 (23.1) | 4 (16.7) | 0 (0.0) | 2 (50.0) |
| c-statics reported | 50 (76.9) | 18 (75.0) | 9 (90.0) | 4 (100.0) |
| c-statics (range) | 0.49 - 0.90 | 0.55 - 0.90 | 0.71 - 0.86 | 0.68 - 0.78 |

*For models with an internal validation

### 2.3. Risk of bias (ROB) and applicability

Of the 242 models, majority of the models had a low ROB for the participant (72%), predictors (94%), and outcome (88%) domains of PROBAST assessment. However, only 4% of all models had low and 85% had high ROB in the analysis domain (Figure 2). The overall ROB was high in 88%, low 2% and unclear in 10% of the models.

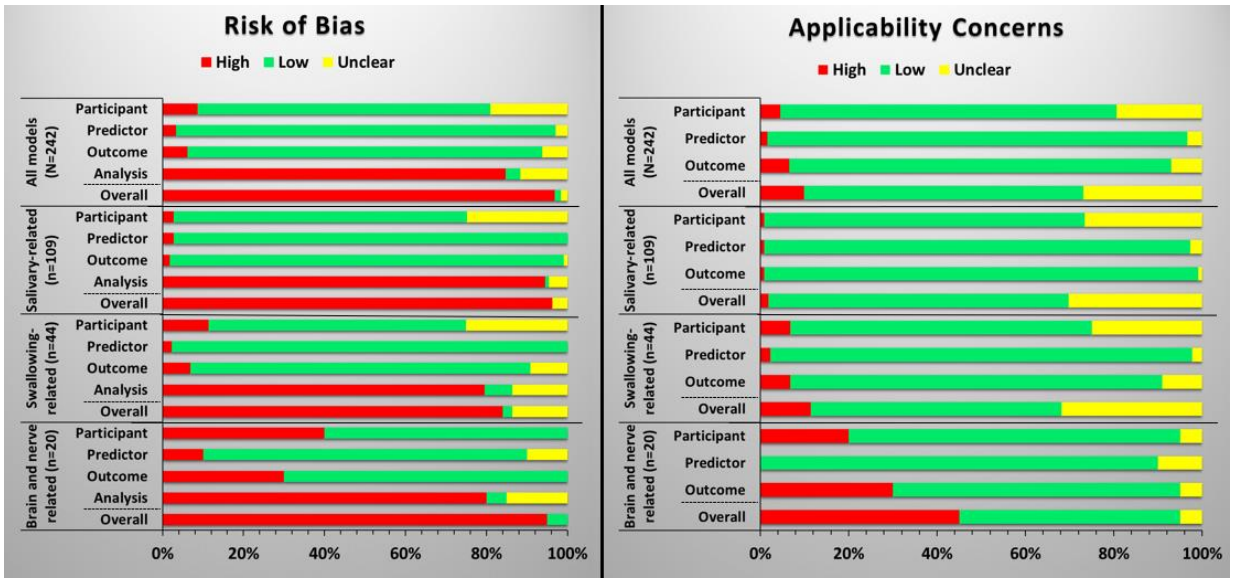



**Figure 2. Risk of bias and applicability concern assessment of all developed models and models for three most modelled outcomes (swallowing-related, salivary-related, and brain and nerve-related) by using PROBAST.**

Similarly, most of the models had low concerns of applicability in the participant (76%), predictors (95%) and outcome (86%) domains according to PROBAST. The overall applicability concern was high in 10%, low in 63% and unclear in 27% of the models (Figure 2). The frequency of the models with a high applicability concern in overall judgement was higher in brain and nerve related outcome (45%) compared with swallowing (11%) and salivary (2%) related outcomes (Figure 2).

## 3. External validation studies

For the 242 models included in model development studies, 64 external validation studies were performed. were externally validated. External validation was most often performed for models that predicted outcomes related to saliva (27%, 29/109), followed by outcomes related to swallowing (41%, 18/44), hypothyroidism (100%, 15/15), brain and nerve (5%, 1/20), and mucosa (7%, 1/14) (Figure 3). No external validation studies were found for models predicting other outcomes.

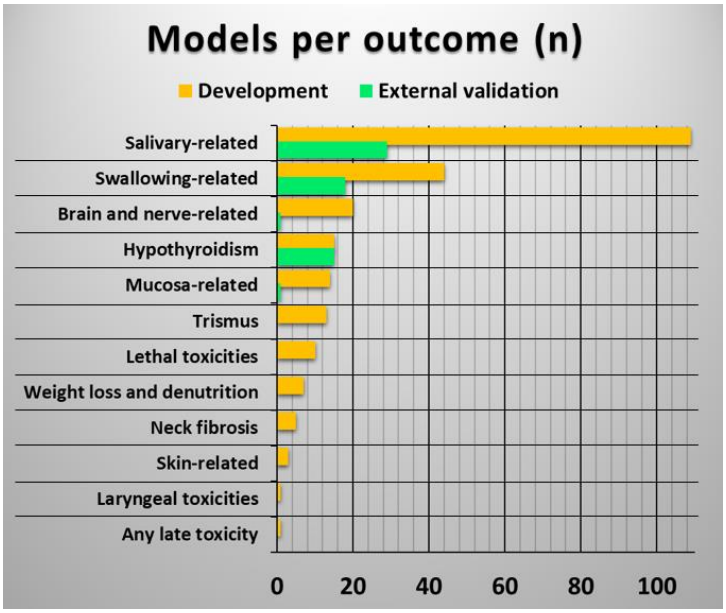



**Figure 3. Number of models per outcome in model development studies and in external validation studies.**

Most models were externally validated in individuals that participated in a randomized trial (n= 29 models, 45%) involving generally <5 centres (n=35 models, 55%) from North America (n = 19, 30%) and Europe (n = 15, 23%). Patient recruitment method was consecutive in 10 models (16%) and not clear in 53 models (83%).

### 3.1. Study inclusion criteria

The inclusion criteria features of the 64 externally validated models and three most validated outcomes among them (salivary-related, swallowing-related, and hypothyroidism) are given in the Table 4.

Of the 64 external validation studies, tumour site was specified in the study inclusion criteria as nasopharynx in 12 studies, oropharynx in 10 studies and tongue 10 studies, while it was multiple tumour sites in 25% and non-specified in 25% of the studies, indicating that approximately half of the external studies were targeted to a specific tumour site. However, histological tumour type was not specified in study inclusion criteria of most models (n= 53, 83%).

The most common RT techniques were IMRT (including VMAT and tomotherapy) (n = 40, 63%) study inclusion criteria. However, neither RT target dose (n = 54, 84%) nor RT fractionation scheme (n=53, 83%) or chemotherapy administration

(n=55, 86%) were specified. Receiving no surgery was an inclusion criterion for in 61 (95%) external validation studies, meaning most external validation studies focused on a head and neck patient population treated with radical RT, mostly using IMRT.

**Table 4. Characteristics of study inclusion criteria for all models and group of toxicities with the highest number of models with an external validation.**

| Variable n (%) | Number of external validation studies | Models for toxicities related to saliva | Models for toxicities related to swallowing | Models for hypothyroidism |
|---|---|---|---|---|
| **Number of models** | 64 (100) | 29 | 18 | 15 |
| **Region** | | | | |
| North America | 19 (29.7) | 1 (3.5) | 8 (44.4) | 8 (53.3) |
| Europe | 15 (23.4) | 3 (10.3) | 7 (38.9) | 5 (33.3) |
| Asia | 5 (7.8) | 0 (0.0) | 3 (16.7) | 2 (13.3) |
| Australia | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Combination | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Not reported | 25 (39.1) | 25 (86.2) | 0 (0.0) | 0 (0.0) |
| **N of centres** | | | | |
| < 5 | 35 (54.7) | 4 (13.8) | 14 (77.8) | 15 (100.0) |
| 5-9 | 4 (6.3) | 0 (0.0) | 4 (22.2) | 0 (0.0) |
| > 10 | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Not reported | 25 (39.1) | 25 (86.2) | 0 (0.0) | 0 (0.0) |
| **Design** | | | | |
| Randomised trial participants | 29 (45.3) | 25 (86.2) | 4 (22.2) | 0 (0.0) |
| Prospective cohort | 16 (25.0) | 3 (10.3) | 5 (27.8) | 6 (40.0) |
| Retrospective cohort | 11 (17.2) | 0 (0.0) | 2 (11.1) | 9 (60.0) |
| Others | 1 (1.6) | 1 (3.4) | 0 (0.0) | 0 (0.0) |
| Unclear | 7 (10.9) | 0 (0.0) | 7 (38.9) | 0 (0.0) |
| **Patient recruitment** | | | | |
| Consecutive | 10 (15.6) | 2 (6.9) | 3 (16.7) | 5 (33.3) |
| Non-consecutive | 1 (1.6) | 1 (3.4) | 0 (0.0) | 0 (0.0) |
| Unclear | 53 (82.8) | 26 (89.7) | 15 (83.3) | 10 (66.7) |
| ***Study inclusion criteria*** | | | | |
| **Tumour site** | | | | |
| Multiple sites | 16 (25.0) | 8 (27.6) | 6 (33.3) | 1 (6.7) |
| Nasopharynx | 12 (18.8) | 10 (34.5) | 0 (0.0) | 2 (13.3) |
| Oropharynx | 10 (15.6) | 0 (0.0) | 0 (0.0) | 10 (66.7) |
| Tongue | 10 (15.6) | 10 (34.5) | 0 (0.0) | 0 (0.0) |
| Non-specified | 16 (25.0) | 1 (3.4) | 12 (66.7) | 2 (13.3) |
| **Cancer type** | | | | |
| Non-specified | 53 (82.8) | 24 (82.8) | 12 (66.7) | 15 (100.0) |
| SCC | 11 (17.2) | 5 (17.2) | 6 (33.3) | 0 (0.0) |
| Others | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| **Cancer stage** | | | | |
| Specified in inclusion criteria | 59 (92.2) | 29 (100.0) | 18 (100.0) | 10 (66.7) |
| Non-specified | 5 (7.8) | 0 (0.0) | 0 (0.0) | 5 (33.3) |
| **Type of radiotherapy technique** | | | | |
| IMRT (incl. VMAT and tomotherapy) | 40 (62.5) | 22 (75.9) | 8 (44.4) | 10 (66.7) |

| | | | | |
|---|---|---|---|---|
| Combination of different techniques | 8 (12.5) | 5 (17.2) | 1 (5.6) | 2 (13.3) |
| Proton therapy | 5 (7.8) | 1 (3.4) | 2 (11.1) | 1 (6.7) |
| 3D conformal | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| SBRT | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Non-specified | 11 (17.2) | 1 (3.4) | 7 (38.9) | 2 (13.3) |
| **Target radiation dose** | | | | |
| Case mix of different target doses | 7 (10.9) | 1 (3.4) | 2 (11.1) | 4 (26.7) |
| ≥ 66Gy | 3 (4.7) | 2 (6.9) | 1 (5.6) | 0 (0.0) |
| < 66 Gy | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Non-specified | 54 (84.4) | 26 (89.7) | 15 (83.3) | 11 (73.3) |
| **Radiotherapy fractionation** | | | | |
| Conventional fractionation | 6 (9.4) | 3 (10.3) | 1 (5.6) | 2 (13.3) |
| Non-conventional fractionation | 2 (3.1) | 0 (0.0) | 2 (11.1) | 0 (0.0) |
| Case mix with different fractionations | 3 (4.7) | 0 (0.0) | 3 (16.7) | 0 (0.0) |
| Non-specified | 53 (82.8) | 26 (89.7) | 12 (66.7) | 13 (86.7) |
| **Type of surgery** | | | | |
| No surgery | 61 (95.3) | 29 (100.0) | 18 (100.0) | 13 (86.7) |
| Case mix | 3 (4.7) | 0 (0.0) | 0 (0.0) | 2 (13.3) |
| Post-surgery | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Non-specified | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| **Chemotherapy** | | | | |
| Yes | 9 (14.1) | 2 (6.9) | 4 (22.2) | 2 (13.3) |
| No chemotherapy | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Non-specified | 55 (85.9) | 27 (93.1) | 14 (77.8) | 13 (86.7) |

### 3.2. Characteristics of models

The characteristics of the 64 models and three most modelled outcomes (salivary-related, swallowing-related, and hypothyroidism) were given in the Table 5.

Models for toxicities related to saliva production (n= 29, 45%) constituted majority of the external validated models, followed by swallowing related complications (n = 18, 28%), and hypothyroidism (n = 15, 23%). Most models were externally validated by either the same author or group (n=37, 58%), rather than a different author or group (n=27, 42%). Eighty-one percent of the external validations was done using the original model with regression coefficients, while in 8% a different predictor was used.

The median age of the participants ranged between 49.5 – 61.0 in 22 models (63%) where reported. The male participants consisted of 70 – 87% of the study population in 34 models (35%), where reported. In 55 (86%) of the external validation studies, the number of participants with the outcome was ≥100.

Calibration of the original model and of an updated for the same existing model were reported in 11 (17%) and 4 (6%) models. The c-statistics without update was reported for 49 models (77%). The c-statistics of the updated model was reported in 8 models (13%) which ranged 0.72 to 0.73.

**Table 5. Model characteristics of the whole models and group of toxicities with the highest number of models (external validation studies).**

| Variable n (%) | Number of external validation studies | Models for toxicities related to saliva | Models for toxicities related to swallowing | Models for hypothyroidism |
|---|---|---|---|---|
| **Number of models** | 64 (100) | 29 | 18 | 15 |
| **Original publication** | | | | |
| Buettner (2012) | 25 (39.1) | 25 (86.2) | - | - |
| Boomsma (2012) | 10 (15.6) | - | - | 10 (66.7) |
| Dean (2018) | 6 (9.4) | - | 6 (33.3) | - |
| Christianen (2012) | 5 (7.8) | - | 5 (27.8) | - |
| Wopken (2014) | 4 (6.2) | - | 4 (22.2) | - |
| Bakhshandeh (2013) | 3 (4.7) | - | - | 3 (20.0) |
| Langendijk (2009) | 2 (3.1) | - | 2 (11.1) | - |
| Ronjom (2013) | 2 (3.1) | - | - | 2 (13.3) |
| Beetz_1 (2012) | 2 (3.1) | 2 (6.9) | - | - |
| Beetz_2 (2012) | 1 (1.6) | 1 (3.4) | - | - |
| Bhide (2012) | 1 (1.6) | - | - | - |
| dean (2018) | 1 (1.6) | - | 1 (5.6) | - |
| Schuette (2019) | 1 (1.6) | - | - | - |
| Tenhunen (2008) | 1 (1.6) | 1 (3.4) | - | - |
| **Validation performed by** | | | | |
| Same author | 34 (53.1) | 27 (93.1) | 7 (38.9) | 0 (0.0) |
| Different author | 27 (42.2) | 1 (3.4) | 9 (50.0) | 15 (100.0) |
| Same group | 3 (4.7) | 1 (3.4) | 2 (11.1) | 0 (0.0) |
| **Outcome** | | | | |
| Salivary-related | 29 (45.3) | 29 (100.0) | | |
| Swallowing-related | 18 (28.1) | | 18 (100.0) | |
| Hypothyroidism | 15 (23.4) | | | 15 (100.0) |
| Brain and nerve-related | 1 (1.6) | | | |
| Mucosa-related | 1 (1.6) | | | |
| **Version of the validated model** | | | | |
| Original model with regression coefficients | 52 (81.2) | 29 (100.0) | 13 (72.2) | 9 (60.0) |
| Different predictor | 5 (7.8) | 0 (0.0) | 0 (0.0) | 5 (33.3) |
| Simplified model based on a risk score | 2 (3.1) | 0 (0.0) | 2 (11.1) | 0 (0.0) |
| All Beta was re-estimated | 1 (1.6) | 0 (0.0) | 1 (5.6) | 0 (0.0) |
| Intercept refitting | 1 (1.6) | 0 (0.0) | 1 (5.6) | 0 (0.0) |
| Slope refitting | 1 (1.6) | 0 (0.0) | 1 (5.6) | 0 (0.0) |
| Updated original model | 1 (1.6) | 0 (0.0) | 0 (0.0) | 1 (6.7) |
| Nomogram | 1 (1.6) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| **N of participant with outcome** | | | | |
| ≥ 100 | 55 (85.9) | 28 (96.5.0) | 18 (100.0) | 8 (53.3) |
| < 100 | 7 (10.9) | 1 (3.5) | 0 (0.0) | 7 (46.7) |
| Not reported | 2 (3.1) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| **Age not reported** | | | | |
| Not reported | 41 (64.0) | 28 (96.6) | 12 (66.7) | 0 (0.0) |
| Range | 49.5 - 61.0 | 60.0-60.0 | 59.0 - 63.0 | 49.5 - 61.0 |
| **Proportion of male** | | | | |
| Not reported | 30 (64.8) | 26 (89.7) | 4 (22.2) | 0 (0.0) |
| Range | 70.0 - 87.0 | 70.0-71.0 | 70.0 - 87.2 | 70.0 - 87.0 |
| **Calibration in original model reported** | 11 (17.2) | 2 (6.9) | 5 (27.8) | 3 (20.0) |

| | | | | |
|---|---|---|---|---|
| **c-statics (without update)** | | | | |
| Not reported | 15 (23.4) | 1 (3.4) | 3 (16.7) | 10 (66.7) |
| Range | 0.64 - 0.91 | 0.19 - 0.96 | 0.68 - 0.95 | 0.64 - 0.91 |
| **Calibration in updated model reported** | 4 (6.2) | 29 (100.0) | 4 (22.2) | 15 (100.0) |
| **c-statics (with update)** | | | | |
| Not reported | 56 (85.9) | 27 (93.1) | 17 (94.4) | 9 (60.0) |
| Range | 0.72 - 0.73 | 0.66 - 0.68 | 0.68 - 0.68 | 0.72 - 0.73 |

### 3.3. Risk of bias (ROB) and applicability

Of the 64 validation studies, most had a low ROB for the predictors (98%) and outcome (70%) domains of the PROBAST tool. However, 47% models for hypothyroidism had a high ROB in outcome domain.

In terms of participants, ROB was high in 19%, low in 33% and unclear in 48% of the models. The ROB for participants was high in 60% of models for hypothyroidism, while it was unclear in 61% and 69% of models for swallowing and for salivary related toxicities.

Like the model development studies, the analysis domain was considered at high risk of bias for most external validation studies (86%). The overall ROB was high in 97%, low 2% and unclear in 2% of the models (Figure 4).

The overall assessment of the models resulted in an unclear applicability concern in most models (59%), whereas it was high in 17% and low in 23% of the models. Applicability concerns in overall judgement were higher in models for hypothyroidism (n=60%), which was caused by the high applicability concern in the participants (60%) and outcome (47%) domains. On the other hand, unclear applicability concern was predominant in models for salivary (72%) and swallowing (67%) related toxicities due to unclear applicability concern in participant domain for both model groups (Figure 4).

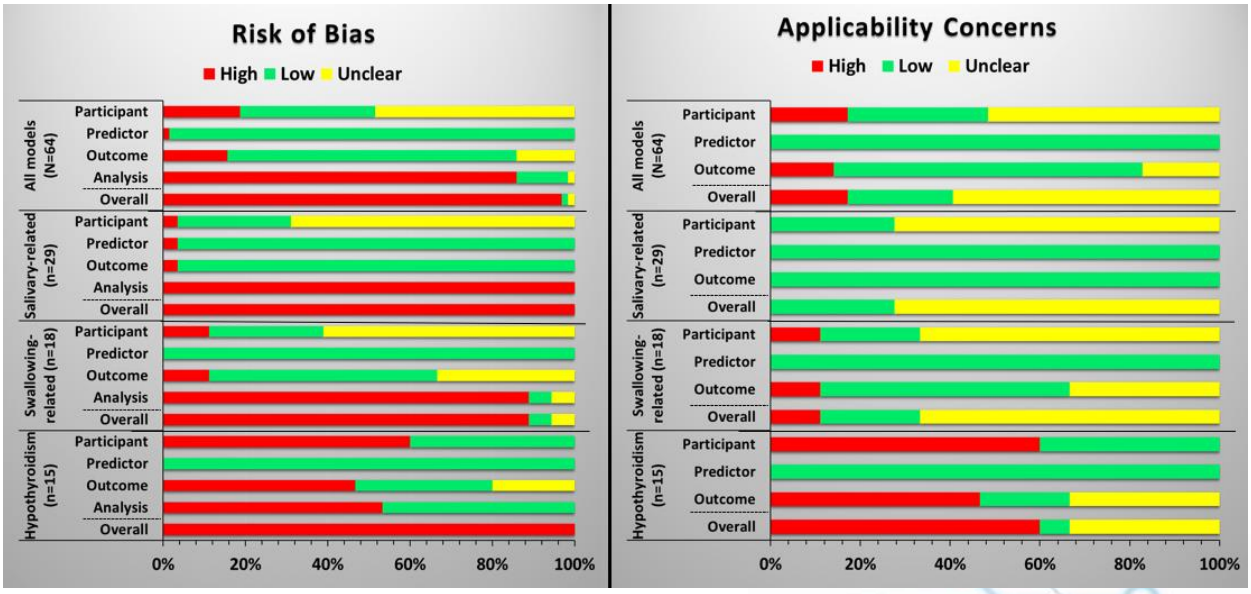



**Figure 4. Risk of bias and applicability concern assessment of all externally validated models and models for three most modelled outcomes (swallowing-related, salivary-related, and hypothyroidism) by using PROBAST.**

## Conclusions

### Implications for practice

We identified 242 models that were developed and 64 external validation studies, mostly for the prediction of toxicities related to saliva and swallowing. The number of external validation studies ranged between 1 to 4 for models with an external validation. In most studies, inclusion criteria were poorly reported, and a proper reporting of the model performance was lacking. Ninety-seven percent of both the model development and external validation studies had an overall high risk of bias according to PROBAST, mainly due to a high risk of bias in the analysis domain, while applicability concerns were mostly low in model development studies (63%) but often unclear in external validation studies (59%). Four models in the development studies had a low risk of bias (3 low and 1 with unclear applicability concern) whereas only 1 external validation study was assessed as having low risk of bias and applicability concern. Therefore, most of the models cannot be used to provide an approximate radiotherapy-induced toxicity prediction but require testing and updating before being applied to new patients.

### Implications for research

Most studies (including the recent studies) were not carried out in accordance with the current standards of prediction model development. Hence, our recommendation is to carefully read the most up to date guidelines and tools, such as CHARMS, TRIPOD and PROBAST at the beginning of study design and follow them at each step of model development and validation to avoid risk of bias and applicability concerns in the future studies.

Our review also demonstrated that there is a great need for external model validation studies. Especially with the increased interest in machine learning techniques for modelling and data obtained from different imaging modalities and 3D radiation dose distribution as a predictor in addition to classical clinical features and 2D dose-volume histogram input, external validation is of utmost important. To provide reliable models that can be used for future patients, we advise to perform an external validation parallel to the model development studies as a first essential step towards future implementation of a (new) prediction model.

## References


Beetz, I., Schilstra, C., Burlage, F. R., Koken, P. W., Doornaert, P., Bijl, H. P., . . . Langendijk, J. A. (2012). 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, 105*(1), 86-93. doi:10.1016/j.radonc.2011.05.010

Cheraghi, S., Nikoofar, A., Bakhshandeh, M., Khoei, S., Farahani, S., Abdollahi, H., & Mahdavi, S. R. (2017). Normal tissue complication probability modeling of radiation-induced sensorineural hearing loss after head-and-neck radiation therapy. *Int J Radiat Biol, 93*(12), 1327-1333. doi:10.1080/09553002.2017.1385872

Christianen, M. E., van der Schaaf, A., van der Laan, H. P., Verdonck-de Leeuw, I. M., Doornaert, P., Chouvalova, O., . . . Langendijk, J. A. (2016). Swallowing sparing intensity modulated radiotherapy (SW-IMRT) in head and neck cancer: Clinical validation according to the model-based approach. *Radiother Oncol, 118*(2), 298-303. doi:10.1016/j.radonc.2015.11.009

Debray, T. P., Damen, J. A., Snell, K. I., Ensor, J., Hooft, L., Reitsma, J. B., . . . Moons, K. G. (2017). A guide to systematic review and meta-analysis of prediction model performance. *BMJ, 356*, i6460. doi:10.1136/bmj.i6460

Emami, B. (2013). Tolerance of Normal Tissue to Therapeutic Radiation. *Reports of Radiotherapy and Oncology, 1*(1), 35-48.

Geersing, G. J., Bouwmeester, W., Zuithoff, P., Spijker, R., Leeflang, M., & Moons, K. G. (2012). Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLoS One, 7*(2), e32844. doi:10.1371/journal.pone.0032844

Halperin, E. C. P., C.A.; Brady,L.W.;. (2008). *Perez and Brady's principles and practice of radiation oncology* (5th ed.). Philadelphia: Lippincott Williams & Wilkins.

Kierkels, R. G., Korevaar, E. W., Steenbakkers, R. J., Janssen, T., van't Veld, A. A., Langendijk, J. A., . . . van der Schaaf, A. (2014). Direct use of multivariable normal tissue complication probability models in treatment plan optimisation for individualised head and neck cancer radiotherapy produces clinically acceptable treatment plans. *Radiother Oncol, 112*(3), 430-436. doi:10.1016/j.radonc.2014.08.020

Langendijk, J. A., Doornaert, P., Verdonck-de Leeuw, I. M., Leemans, C. R., Aaronson, N. K., & Slotman, B. J. (2008). Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. *J Clin Oncol, 26*(22), 3770-3776. doi:10.1200/JCO.2007.14.6647

Langendijk, J. A., Lambin, P., De Ruysscher, D., Widder, J., Bos, M., & Verheij, M. (2013). Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. *Radiother Oncol, 107*(3), 267-273. doi:10.1016/j.radonc.2013.05.007

Luo, R., Wu, V. W. C., He, B., Gao, X., Xu, Z., Wang, D., . . . Lin, Z. (2018). Development of a normal tissue complication probability (NTCP) model for radiation-induced hypothyroidism in nasopharyngeal carcinoma patients. *BMC Cancer, 18*(1), 575. doi:10.1186/s12885-018-4348-z

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ, 339*, b2535. doi:10.1136/bmj.b2535

Moons, K. G., de Groot, J. A., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D. G., . . . Collins, G. S. (2014). Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med, 11*(10), e1001744. doi:10.1371/journal.pmed.1001744

Moons, K. G. M., Wolff, R. F., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., . . . Mallett, S. (2019). PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med, 170*(1), W1-W33. doi:10.7326/M18-1377

Takada, T., Damen, J. A. A. G., Tambas, M., Spijker, R., Steenbakkers, R. J. H. M., Sharabiani, M., . . . Schuit, E. (2021). Prognostic models for radiation-induced complications a+er radiotherapy in head and neck cancer patients (Protocol). *Cochrane Database of Systematic Reviews, Issue 4. Art. No.: CD014745.* doi:10.1002/14651858.CD014745.

Widder, J., van der Schaaf, A., Lambin, P., Marijnen, C. A. M., Pignol, J. P., Rasch, C. R., . . . Langendijk, J. A. (2016). The Quest for Evidence for Proton Therapy: Model-Based Approach and Precision Medicine. *Int J Radiat Oncol Biol Phys, 95*(1), 30-36. doi:10.1016/j.ijrobp.2015.10.004

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., . . . Groupdagger, P. (2019). PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med, 170*(1), 51-58. doi:10.7326/M18-1376

Wopken, K., Bijl, H. P., van der Schaaf, A., van der Laan, H. P., Chouvalova, O., Steenbakkers, R. J., . . . Langendijk, J. A. (2014). Development of a multivariable normal tissue complication probability (NTCP) model for tube feeding dependence after curative radiotherapy/chemo-radiotherapy in head and neck cancer. *Radiother Oncol, 113*(1), 95-101. doi:10.1016/j.radonc.2014.09.013