Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Oral Oncol. Author manuscript; available in PMC 2025 April 01.

Published in final edited form as:

Oral Oncol. 2025 April ; 163: 107224. doi:10.1016/j.oraloncology.2025.107224.

Evaluation of a comprehensive set of normal tissue complication probability models for patients with head and neck cancer in an international cohort

Suzanne P.M. de Vette[a,*,1], **Maria I. van Rijn-Dekker**[a,1], **Lisa Van den Bosch**[a], **Kylie Keijzer**[a,b], **Hendrike Neh**[a], **Hung Chu**[a,2], **Yan Li**[a], **Mark L. Frederiks**[a], **Hans Paul van der Laan**[a], **Jolien Heukelom**[c], **Peter van Luijk**[a], **Arjen van der Schaaf**[a], **Roel J.H.M. Steenbakkers**[a], **Nanna M. Sijtsema**[a], **Katherine A. Hutcheson**[d,e], **Clifton D. Fuller**[d], **Johannes A. Langendijk**[a], **Amy C. Moreno**[d], **Lisanne V. van Dijk**[a]

[a]Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

[b]Department of Oncology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

[c]Department of Radiation Oncology, MAASTRO, University of Maastricht, Maastricht, the Netherlands

[d]Department of Radiation Oncology, the University of Texas MD Anderson Cancer Center, Houston, TX, USA

[e]Department of Head and Neck Surgery, the University of Texas MD Anderson Cancer Center, Houston, TX, USA

Abstract

This is an open access article under the CC BY-NC license (https://creativecommons.org/licenses/by-nc/4.0/).

*Corresponding author: s.p.m.de.vette@umcg.nl (S.P.M. de Vette).

[1]Maria I. van Rijn-Dekker and Suzanne P.M. de Vette made equal contributions to this study.

[2]Present address: Center for Information Technology, University of Groningen, Groningen, the Netherlands.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Suzanne P.M. de Vette: Data curation, Formal analysis, Investigation, Methodology, Visualization, Software, Writing – original draft. **Maria I. van Rijn-Dekker:** Formal analysis, Investigation, Methodology, Visualization, Software, Writing – original draft. **Lisa Van den Bosch:** Conceptualization, Software, Writing – review & editing. **Kylie Keijzer:** Investigation, Formal analysis, Writing – review & editing. **Hendrike Neh:** Investigation, Formal analysis, Writing – review & editing. **Hung Chu:** Investigation, Formal analysis, Writing – review & editing. **Yan Li:** Investigation, Formal analysis, Writing – review & editing. **Mark L. Frederiks:** Investigation, Formal analysis, Writing – review & editing. **Hans Paul van der Laan:** Investigation, Formal analysis, Writing – review & editing. **Jolien Heukelom:** Investigation, Formal analysis, Writing – review & editing. **Peter van Luijk:** Supervision, Writing – review & editing. **Arjen van der Schaaf:** Supervision, Writing – review & editing. **Roel J.H.M. Steenbakkers:** Supervision, Writing – review & editing. **Nanna M. Sijtsema:** Supervision, Writing – review & editing. **Katherine A. Hutcheson:** Supervision, Data curation, Writing – review & editing. **Clifton D. Fuller:** Supervision, Writing – review & editing. **Johannes A. Langendijk:** Conceptualization, Supervision, Writing – review & editing. **Amy C. Moreno:** Data curation, Methodology, Supervision, Formal analysis, Investigation, Writing – review & editing. **Lisanne V. van Dijk:** Conceptualization, Data curation, Methodology, Software, Supervision, Writing – review & editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.oraloncology.2025.107224.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Background/purpose: Normal tissue complication probability (NTCP) models can be used to guide radiation therapy (RT) decisions by estimating side-effect risks pretreatment to minimize (late) side-effects. Recently, a comprehensive individual toxicity risk (CITOR) profile of NTCP models addressing common side-effects in head and neck cancer (HNC) patients was developed. This study investigates the generalizability of these models in an international setting, with different treatment approaches and side-effect assessments, promoting their integration into more widespread clinical practice.

Materials/methods: From a prospective registry study, 407 HNC patients were included who were treated with definitive RT with or without systemic therapy between 2015 and 2022. NTCP models predicting dysphagia, aspiration, xerostomia, sticky saliva, taste loss, speech problems, oral pain, and fatigue at 6 and 12 months after RT were evaluated. All side-effects were patient-rated using the MDASI-HN, except dysphagia which was reported by clinicians using the PSS-HN diet normalcy score. Model performance was appraised by discrimination (area under the curve [AUC]) and calibration.

Results: CITOR models showed moderate-to-high performance in this cohort (mean AUC = 0.67[range = 0.55–0.80], moderate-to-good calibration). NTCP models for dysphagia, xerostomia, sticky saliva, and fatigue were the top performing models. Models for aspiration, taste loss and speech problems performed moderately well, which was partly explained by lower incidences.

Conclusion: Despite differences between the CITOR development and this evaluation cohort, including use of different side-effect scoring systems, most models exhibited moderate-to-high performance. This demonstrated that the dose–effect relations were generalizable. Therefore, this study supports further integration of these NTCP models in clinical practice.

Keywords

Normal Tissue Complication Probability; Head and Neck cancer; Independent evaluation; Comprehensive Individual Toxicity Risk; Radiation therapy; Late toxicity

Introduction

Head and neck cancer (HNC) patients often experience severe side-effects during and after radiation therapy (RT) treatment, heavily impacting patients' health and quality of life [1–3]. Developing side-effects is associated with the radiation dose delivered to organs-at-risk (OARs)[4–7]. For instance, radiation dose to the parotid glands and submandibular glands has been associated with the risk of xerostomia [7–9], while higher doses to the pharyngeal constrictor muscles (PCMs) and larynx increase the risk of dysphagia and aspiration [6]. Normal tissue complication probability (NTCP) models estimate side-effect development risk with OAR doses as predictors. Historically, NTCP models were Lyman-Kutcher-Burman (LKB) models, using one dose parameter of a single OAR [10,11]. These models have evolved from univariable LKB to multivariable logistic regression NTCP models, which can describe the dose–effect relations, together with other potential risk factors (i.e. baseline complaints, age) [4,5]. By providing patient-specific side-effect risks, NTCP models can guide individual treatment optimization to reduce dose to specified OARs to minimize side-effects [10,11]. The most optimal treatment plan or radiation technique can be selected by comparing the predicted risk of side-effects (i.e., proton versus photon

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

therapy, as per clinical standards in the Netherlands) [12,13]. For NTCP models to be an usable tool for clinical practice globally, careful evaluation in different treatment settings is necessary [12].

Multiple NTCP models exist for HNC patients, predicting side-effects such as xerostomia [9], dysphagia [14], and taste loss [15]. However, using each model separately can lead to dose redistribution, mitigating one complication but increasing risks for others [16]. With several models available, selecting the right ones for treatment optimization is challenging. To address this, Van den Bosch et al. developed the comprehensive individual toxicity risk (CITOR) profile, predicting a wide range of side effects across six domains: swallowing, salivary, mucosal, speech, pain and general [17]. They developed a comprehensive set of NTCP models for 22 common side-effects at 10 time points in 750 HNC patients, and subsequently externally validated them on multi-institutional data [17,18], achieving the highest evidence level (TRIPOD level 4) [19]. However, implementing these NTCP models globally requires evaluation in an international setting, as different treatment approaches and side-effect scoring systems may affect the NTCP models' generalizability [20].

Therefore, this study aimed to evaluate the performance of the CITOR models predicting *dysphagia*, *aspiration*, *xerostomia*, *sticky saliva*, *taste loss*, *speech problems*, *oral pain*, and *fatigue* at 6 and 12 months post-RT in HNC patients treated at the University of Texas MD Anderson Cancer Center (MDACC).

Materials and methods

Study population

Patients with oropharyngeal cancer (OPC) or an unknown primary HNC tumor treated with definitive RT with/without concurrent platinum-based chemotherapy (OPC: 66–70 Gy in 30–33 fractions; unknown primary: 50–57 Gy in 33 fractions) with curative intent between February 2015 and January 2020 were included. All patients participated in a prospective registry study at MDACC, approved by the Institutional Review Board (IRB) (PA14–0947 data collection, PA11–0809 retrospective analysis). Exclusion criteria were primary tumor surgery, neck dissection, induction chemotherapy, recurrent disease, distant metastasis, or age under 18. Dose reduction to the parotid glands, larynx, and esophagus was prioritized (Table A.1, Supplementary materials).

Dose-volume histogram parameters were extracted from radiation dose distributions and OAR segmentations that were used for treatment optimization. Segmentations were generated with the Atlas-based auto-segmentation (ABAS) tool [21] in consensus with the international guidelines [22]. Missing OARs were supplemented with deep learning contouring (DLC) [23]. For patients who had both DLCs and ABAS contours, the mean dose to the OARs was evaluated to identify outliers. The contours of these outliers were visually assessed.

Side-effect assessment

The following side-effects, collected using a prospective registry study, were evaluated at 3–6 and at 12 months after RT: dysphagia and aspiration (swallowing domain), xerostomia,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

sticky saliva, and taste loss (salivary domain), speech problems (speech domain), oral pain (pain domain), and fatigue (general domain). Dysphagia was assessed using the physician-rated diet normalcy score from the Performance Status Scale for Head and Neck cancer (PSS-HN) ranging from 0 (non-oral feeding) to 100 (full diet) [24]. All other side-effects were patient-rated using the MD Anderson Symptom Inventory Head and Neck (MDASI-HN) ranging from 0 (no symptoms) to 10 (symptom as bad as you can imagine) [25]. Scores were dichotomized per below under "Side-effect scoring system comparison". Missing 12-month scores were manually imputed if the 3–6 and 18-month scores fell into the same dichotomized category. If the scores did not fall into the same dichotomized category, these patients were excluded for analysis.

NTCP models

This study evaluated 16 NTCP models predicting moderate-to-severe side-effects at 6 (M6) and 12 months after RT (M12) [17]. Model predictors included a combination of OAR mean doses (D_{mean}) and clinical variables (i.e., baseline complaints or age). For more details see Fig. 1, the logistic regression model coefficients in Table B.1 or Van den Bosch et al. [17], and the NTCP formulas in Supplement B.

Side-effect scoring system comparison

Side-effect assessments at MDACC differed from the CITOR model development study, though both assess the same symptom constructs. MDACC used the MDASI-HN and PSS-HN, while the CITOR study used the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Core 30/Head & Neck 35 (QLQ-C30/H&N35) [26] and Common Terminology Criteria for Adverse Events (CTCAE) [27]. Specific scoring systems are detailed in Table C.1.

To assess side-effect scoring systems correlation, a prospective cross-sectional study (NCT02435576) at the University Medical Center Groningen (UMCG) from November 2023 to February 2024 involved HNC patients completing the MDASI-HN and EORTC QLQ-C30/H&N35 questionnaires during or after RT, while physicians reported CTCAE and PSS-HN. Spearman's rho (ρ) was used to evaluate the correlation between assessments. Optimal thresholds for MDASI-HN (0–10 scale) and PSS-HN (0–100 scale) corresponding to EORTC QLQ-C30/H&N35 and CTCAE grade ≥ 2 side effects were determined using Youden's J index, sensitivity, specificity, and Cohen's kappa (κ). The Youden's J index, which balances sensitivity and specificity, was leading. If these results were inconclusive, Cohen's kappa guided the decision for an optimal threshold.

Statistical analysis

Differences in baseline demographic and treatment characteristics between the MDACC cohort and the previously published CITOR development cohort were assessed using One-way analysis of variance, Kruskal-Wallis, χ^2, and Fisher's exact testing.

NTCP model performance metrics included: 1) area under the receiver operating characteristic curve (AUC) for discrimination capacity, 2) calibration performance with plots, 3) Nagelkerke R-squared (R^2) for explained variance. Performance was considered

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

robust with moderate calibration [28] and when discrimination was comparable to the original validation cohort, while considering the differences in used side-effect scoring systems. Model performance was tested on the full MDACC cohort and specific subsets (ABAS delineation only, DLC delineation only, and OPC patients only).

All statistical analyses were performed in R, version 4.1.3 [29].

Results

Study population

In total 407 HNC patients were included (Table 1), but patient numbers per NTCP model differed due to side-effect data availability (Fig. D.1, Table E.1). Most patients had OPC (91 %) and 9 % had an unknown primary tumor. The cohort had a high rate of HPV positive tumors (74 %), Tis-T2 tumors (73 %) and positive lymph nodes (92 %). Most were treated with volumetric modulated arc therapy (VMAT, 57 %), followed by intensity modulated proton therapy (18 %), intensity modulated radiotherapy (IMRT, 17 %) and a VMAT-IMRT combination (8 %). In contrast, the CITOR development cohort included HNC patients with varying tumor sites (i.e., oral cavity, pharynx, larynx), who were treated with 3D-Conformal radiotherapy (11 %), VMAT (14 %), IMRT (73 %), or a VMAT-IMRT combination (2 %), leading to significant differences in OAR radiation doses between the cohorts (Table 1, Fig. F.1) [17]. ABAS OAR delineations were available for 97 %, except for buccal mucosa and external contours, which were all supplemented with DLC (Table E.2).

Side-effect scoring system comparison

In the UMCG study comparing the MDASI-HN/PSS-HN with the EORTC/CTCAE scores, 318–846 paired assessments of 512 HNC patients were analyzed (Table G.1). Significant correlations ($p < 0.001$) were found with strong correlations for almost all side-effects (ρ = 0.70–0.85), except for aspiration ($\rho = 0.61$), sticky saliva ($\rho = 0.64$), and oral pain ($\rho = 0.52$) (Table 2, Fig. H.1).

The most optimal PSS-HN cut-off value corresponding to grade ≥ 2 CTCAE dysphagia was ≤ 50. MDASI-HN cut-offs corresponding to moderate-to-severe EORTC were: ≥5 for aspiration and speech problems, ≥4 for xerostomia, sticky saliva, taste loss and fatigue, and ≥ 2 for oral pain (Table 2). After applying these cut-off values, xerostomia and taste loss showed the highest incidence, while aspiration and speech problems were less frequent (Fig. 1). Compared to the current MDACC cohort, the CITOR development cohort had higher incidences for dysphagia, aspiration, sticky saliva, speech problems, and fatigue (Fig. 2A–B), which were partly explained by baseline differences (Fig. 2C–D).

Performance of NTCP models in MDACC cohort

In the swallowing domain, the *dysphagia* NTCP models overestimated the risk of developing the side-effect, especially at M12 as seen by the difference between the regression and identity line in Fig. 3A–B and the calibration slope of less than 1, with AUCs of AUC_{M6} = 0.66[confidence interval = 0.60–0.73]/AUC_{M12} = 0.64[0.55–0.74] (Table 3). *Aspiration* NTCP models also overestimated the risk (Fig. 3C–D), with AUC_{M6} = 0.72[0.62–0.83]/

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

AUC_{M12} = 0.62[0.49–0.76] (Table 3). For both side-effects, M6 NTCP models performed better than M12 NTCP models.

In the salivary domain, the *xerostomia* NTCP models predicted risks close to the identity line (Fig. 3E–F) and AUCs of AUC_{M6} = 0.63 [0.56–0.69]/AUC_{M12} = 0.69[0.63–0.75] (Table 3). The *sticky saliva* NTCP model at M6 predicted risks close to the identity line, whereas the model at M12 overestimated the risk (Fig. 3G–H), with AUC_{M6} = 0.63 [0.56–0.71]/AUC_{M12} = 0.70[0.62–0.78] (Table 3). Lastly, the *taste loss* NTCP models showed an underestimation of the risk (Fig. 3I–J), with AUCs of AUC_{M6} = 0.55[0.49–0.61]/AUC_{M12} = 0.61[0.54–0.67] (Table 3).

In the speech domain, NTCP models overestimated the risk for *speech problems* (Fig. 3K–L), with AUCs of AUC_{M6} = 0.80[0.68–0.93]/AUC_{M12} = 0.75[0.59–0.90] (Table 3). In the pain domain, the *oral pain* NTCP models underestimated the risk (Fig. 3M–N), with AUCs of AUC_{M6} = 0.66[0.59–0.73]/AUC_{M12} = 0.60[0.51–0.69] (Table 3). In the general domain, the predictions of the *fatigue* NTCP model approached the identity line at M6 and showed an overestimation of the predicted risk at M12 (Fig. 3O–P), with AUCs of AUC_{M6} = 0.73[0.67–0.80]/AUC_{M12} = 0.68[0.59–0.76] (Table 3).

Performance comparisons with the CITOR cohorts and subsets of the MDACC cohort are presented in Table I1–8.

Discussion

The NTCP models showed robust and clinically usable performance at MDACC, despite the major differences in side-effect assessment, demographics, and treatment compared to the model development cohort. Side-effect assessment differences were pursued to be mitigated by identifying the EORTC and CTCAE equivalent cut-off values for the MDASI-HN and PSS-HN scores. Differences in demographics (i.e., HPV prevalence, tumor location, and tumor stage, Table 1), and in treatment approach (i.e., more recent treatment period, different fractionation schedules, radiation optimization and techniques, Table A.1) remained, translating to differences in radiation doses to OARs (Fig. F.1). Despite these differences, most dose–response relations in the NTCP models were generalizable, indicating their operational robustness across cohorts and adequate prediction performance (mean AUC = 0.67 [range = 0.55–0.80], generally with little predicted risk overestimation [Fig. 3]). The current study supports the CITOR NTCP models' potential for integration in clinical practice to predict radiation-induced side-effect risks, (automatically) optimize treatment plans, and/or provide evidence-based selection criteria in patients for less available and costly treatment techniques (adaptive radiotherapy, proton therapy), as has been demonstrated pre-clinically [10–12,30–36]. Specifically, institutes in Denmark and the Netherlands clinically use the *dysphagia* and *xerostomia* NTCP models in the model-based selection of HNC patients for proton therapy [13,37]. The generalizability of the *dysphagia* and *xerostomia* models was shown in this current study.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Performance of models: Close-up per domain

In the swallowing domain, the *dysphagia* models demonstrated generalizability as predicted risks were consistent with observed side-effect incidence (Fig. 3A–B). Moreover, a comparable dose–response relation of mean dose to the PCMs and the oral cavity with dysphagia was found at M06 (Table J.1) [6,38]. Nevertheless, performance (AUC_{M6} = 0.66/AUC_{M12} = 0.64) in the current USA cohort was lower than in the published Dutch validation cohort (AUC_{M6} = 0.75/AUC_{M12} = 0.77, Table I.1). A possible explanation could be the more proactive swallowing therapy and reduced use of prophylactic feeding tubes at MDACC, resulting in less dysphagia at M6 and M12 (Fig. 2)[39]. Furthermore, no patients scored grade ≥ 3 dysphagia at baseline, which reflected a better functioning, mostly HPV-associated OPC subtype in the MDACC cohort and made one of the model parameters redundant. *Aspiration* models performed similarly or better than in the CITOR validation cohort, especially in discrimination (AUC_{M6} = 0.72/AUC_{M12} = 0.62 versus AUC_{M6} = 0.61/AUC_{M12} = 0.63, Table I.2). However, the small number of events in the MDACC cohort (19 at M6, 16 at M12) and moderate Spearman's ρ between the assessments (ρ = 0.61) warrant caution. This does not necessarily mean the model is flawed or nonrobust, but rather that the cohort in which the model was evaluated may not be optimal for determining its predictive performance for patients who develop aspiration. Enhancing the swallowing domain models could involve including doses to additional OARs involved in a correct swallowing maneuver, with principal component analysis used to manage multicollinearity among these variables[38,40]. The models in the swallowing domain, especially the dysphagia models, could help distinguish which patients could benefit from swallowing exercises.

In the salivary domain, the NTCP models predicting *xerostomia* and *sticky saliva* were well-calibrated (Fig. 3C–F) and demonstrated good discrimination, especially at M12 (*xerostomia* AUC_{M6} = 0.63/AUC_{M12} = 0.69, *sticky saliva* AUC_{M6} = 0.63/AUC_{M12} = 0.70), despite moderate Spearman's ρ for sticky saliva assessments (ρ = 0.64). The current study showed that the dose–response relations between the parotid and submandibular gland dose and xerostomia were consistent (Table J.1) [7–9]. However, xerostomia risk was underestimated in the MDACC cohort, especially at M6 (Fig. 3E), possibly due to earlier assessments (i. e., 3–6 months versus 6 months, incidences in Fig. 2), which allowed less time for recovery [8]. Xerostomia and sticky saliva models could help identify which patients would benefit from more advanced treatment techniques, such as transplantation of salivary gland stem cells to prevent post-radiotherapy xerostomia [41]. Although taste loss assessments were strongly correlated (ρ = 0.84), the *taste loss* NTCP models showed insufficient performance (AUC_{M6} = 0.55/AUC_{M12} = 0.61). Taste loss is a complex side-effect that is difficult to predict, for instance due to the unresolved consensus on its mechanism or to its interconnectivity with xerostomia and the sense of smell [42–44]. This was further exemplified by the additional univariable analysis that showed no significant association between taste loss at 6 months and mean oral cavity dose, mean parotid gland dose, or age (Table J.1), contributing to the insufficient performance of this model in this cohort. Deep learning techniques might improve prediction accuracy for taste loss by capturing the complex nature of taste loss, as was previously shown for xerostomia [45].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The *speech problems* models showed good discrimination (AUC_{M6} = 0.80/AUC_{M12} = 0.74) and one of the best goodness-of-fit (R^2_{M6} = 0.28/R^2_{M12} = 0.13), but inconclusive calibration (Fig. 3K–L). The incidence of speech problems was low in this cohort (Fig. 2), which was likely due to the predominance of OPC patients in the MDACC cohort. This could explain the inconclusive calibration compared to the CITOR validation cohort (Table I.6), as OPC patients generally receive less dose to the larynx, and speech problems are more prevalent in patients with laryngeal irradiation[6]. This effect was underlined in a subset of only OPC patients: after excluding patients with an unknown primary tumor, who often received higher doses to the larynx, the model performance dropped (Table I.6). Furthermore, mean doses to the supraglottic larynx or the oral cavity were not significantly associated with speech problems in the complete cohort (Table J.1), suggesting that there might be other causes for patients to develop speech problems within this cohort.

In the pain domain, the *oral pain* NTCP models performed worse in the MDACC cohort compared to the CITOR validation cohort (AUC_{M6} = 0.66/AUC_{M12} = 0.60 versus AUC_{M6} = 0.75/AUC_{M12} = 0.64, Table I.7). This was partly explained by differences in pain assessment (general pain versus oral pain, ρ = 0.52), and lower incidence rates in MDACC, possibly influenced by tumor location differences (i.e., no oral cavity tumors at MDACC) and more aggressive pain management in MDACC. Pain management significantly affects patients' pain perception and reporting in side-effect assessments [46]. However, details on pain mitigation strategies were not included in the model or accounted for in the side-effect evaluations. A possible usage of models predicting pain would be to guide such pain management strategies.

In the general domain, the *fatigue* models demonstrated better performance in the MDACC cohort than in the CITOR validation cohort (AUCM6 = 0.73/AUCM12 = 0.68 versus AUCM6 = 0.70/AUCM12 = 0.61, Table I.8). Differences in rehabilitation approaches and their varying impact[47] between centers may have influenced incidence rates and model performance. Furthermore, including thyroid function as a predictor might improve fatigue predictions, as hypothyroidism is associated with post-treatment fatigue [48]. We pursued to evaluate NTCP models predicting *nausea and vomiting*, but the cut-off value based on the side-effect scoring comparison (i.e., ≥2) resulted in too few events in the MDACC cohort for a valid model evaluation, leading to exclusion of this endpoint.

Differences in OAR delineations arose from using different protocols (ABAS [21] at MDACC cohort versus DLCs [23] in CITOR cohorts). To assess the impact, model performance was evaluated on subsets with OARs delineated using ABAS only and DLC only. While these methods caused dose differences in OAR with high proximity to the tumor, an effect more pronounced in the oral cavity and the PCMs (Fig. K.1), the impact on the NTCP models' performance was limited (Table I.1–8). This suggests that both contouring methods based on the guidelines by Brouwer et al. [22] can be used for these NTCP models.

Previous validation studies

To our knowledge, only the CITOR dysphagia model was validated in a different treatment setting by Kalendralis et al. (2022), demonstrating better performance than in our cohort

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(AUC = 0.80[0.75–0.85] vs 0.66 [0.60–0.73]) [49]. After performing the closed-testing procedure, the revised model achieved an AUC of 0.83[0.75–0.85] [49]. This validation was conducted in a Dutch center on a cohort treated with both photon and proton therapy, which was otherwise comparable to the development cohort. Validation studies of similar dysphagia models based on the superior PCM and supraglottic larynx [14] or oral cavity [13] reported AUCs of 0.63–0.75, aligning with our results [13,50–53]. Calibration results were similar. The xerostomia model by Beetz et al. (2012) was validated multiple times with similar performance to the CITOR xerostomia model in our cohort (AUC = 0.70–0.74) [9,13,53].This study appears to be the first large-scale evaluation of NTCP models across different side-effect assessments, providing a template for centers considering the integration of NTCP models using different side-effect assessments in clinical practice.

Dealing with differences in side-effect assessment

The distinct differences – both phrasing and scaling – between the MDACC questionnaires and EORTC/CTCAE scoring in the NTCP model development cohort (Table C.1), could have compromised the results' interpretation. Hence, an additional prospective study as part of this work was conducted to compare the side-effect scoring systems and determine optimal cut-off values for MDACC outcomes. For most, strong correlations were observed between side-effect assessments, excepting *aspiration*, *sticky saliva*, and *oral pain* (Table 2). Where this can be explained for aspiration by its low incidence (12/442), this was likely for *sticky saliva* and *oral pain* a result from distinct differences in phrasing between assessments (Table C.1). The optimal cut-off values were around ≥ 4 for the MDASI-HN and ≤ 50 for the PSS-HN diet normalcy score (Table 2), corresponding to EORTC moderate-to-severe and CTCAE grade ≥ 2 classification. This aligns with MDACC's empirical experience for clinical intervention. To our knowledge, earlier cut-off values were based on quality of life scales, instead of the same side-effect constructs [25,54].

Limitations

This study faced limitations typical of prospective data registry studies, including missing follow-up data due to patients loss to follow-up and non-compliance (Table E.1) and a limited number of events for some side-effects (Table 3). Consequently, the recommendation to have at least 100 (non–)events for external validation of prognostic models [55] was not always met. Despite this, the number of included patients and events in the current study was comparable or higher than in other published model validation studies (median number of patients/events of 326/60 versus 149/40)[56]. Additionally, the current study included HNC patients with an unknown primary tumor although these patients were not represented in the development cohort. However, the model performance was similar after excluding these patients (Table I.1–8). Furthermore, in the development cohort, toxicities were consistently assessed at 6 months after treatment, but in the current study, they were evaluated between 3–6 months after treatment (as is standard practice in the MDACC), possibly containing the transition from acute to late toxicities. Lastly, characteristics of the US patient cohort (HPV status, tumor stage, treatment modality) differed significantly from the development cohort, which can be seen as a limitation, but also proves the generalizability of these models.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Conclusion

Most evaluated NTCP models from the CITOR profile exhibited robust performance in a very distinct cohort with different side-effect assessments, indicating consistent dose–response relations between the current MDACC cohort and the published development and validation cohorts, especially of the models predicting dysphagia, xerostomia, sticky saliva, and fatigue. Therefore, this study supports the generalizability of these NTCP models. Consequently, the results were promising for the further integration of these NTCP models in clinical practice in an international setting, enabling personalized treatment to minimize side-effects.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Direct infrastructure support was provided by the multidisciplinary Stiefel Oropharyngeal Research Fund of the University of Texas MD Anderson Cancer Center, Charles and Daneen Stiefel Center for Head and Neck Cancer, National Institutes of Health (NIH) National Cancer Institute (NCI) MD Anderson Cancer Center Support Grant (CCSG) Image-Driven Biologically-Informed Therapy Program (P30CA016672).

Funding Statement

[Dr. van Dijk received funding related to this project by KWF Dutch Cancer Society via a Young Investigator Grant (KWF-13529). Moreover, Dr. van Dijk, received/receives salary support related to this project from NWO ZonMw via the VENI (NWO-09150162010173). Y. Li received/receives funding and salary support from the China Scholarship Council grant (202008110193). P. van Luijk and R.J.H.M. Steenbakkers received grants from the Dutch Cancer Society. K. Hutcheson received grant support from Stiefel Oropharyngeal Research Fund (Patient Reported Outcomes/Function Core, PA14–0947), Patient-Centered Outcomes Research Institute (PCS-1609–36195), Department of Defense (W81XWH-20-PRCRP-BHSA), NIH/NCI (1R21CA273984–01 / FP00014732), NIH/NCI (FP00013681), NIH/NCI (PAR-20–060 U01), ATOS PATH funding, Atos Medical. C.D. Fuller received direct grant support from NIDCR (R01DE028290) and NCI (R01CA258827, R01CA257814). J.A. Langendijk received grants from the European Union, Dutch Cancer Society and IBA and his department has collaborative research contracts with financial support with IBA, RaySearch, Elekta, Mirada, Siemens. A.C. Moreno received direct salary support from the National Institute of Dental and Craniofacial Research (NIDCR) Mentored Career Development Award to Promote Diversity in the Dental, Oral and Craniofacial Workforce Award (K01DE030524).].

References

[1]. Sciubba JJ, Goldenberg D. Oral complications of radiotherapy. Lancet Oncol 2006; 7:175–83. 10.1016/S1470-2045(06)70580-0. [PubMed: 16455482]

[2]. Hunter KU, Schipper M, Feng FY, Lyden T, Haxer M, Murdoch-Kinch C-A, et al. Toxicities Affecting Quality of Life After Chemo-IMRT of Oropharyngeal Cancer: Prospective Study of Patient-Reported, Observer-Rated, and Objective Outcomes. International Journal of Radiation Oncology*Biology*Physics 2013;85:935–40. 10.1016/j.ijrobp.2012.08.030.

[3]. van der Laan HP, Van den Bosch L, Schuit E, Steenbakkers RJHM, van der Schaaf A, Langendijk JA. Impact of radiation-induced toxicities on quality of life of patients treated for head and neck cancer. Radiother Oncol 2021;160:47–53. 10.1016/j.radonc.2021.04.011. [PubMed: 33892023]

[4]. Brodin NP, Kabarriti R, Garg MK, Guha C, Tomé WA. Systematic Review of Normal Tissue Complication Models Relevant to Standard Fractionation Radiation Therapy of the Head and Neck Region Published After the QUANTEC Reports. International Journal of Radiation Oncology*Biology*Physics 2018;100:391–407. 10.1016/j.ijrobp.2017.09.041.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[5]. Stieb S, Lee A, van Dijk LV, Frank S, Fuller CD, Blanchard P. NTCP Modeling of Late Effects for Head and Neck Cancer: A Systematic Review. Int J Part Ther 2021;8:95–107. 10.14338/20-00092. [PubMed: 34285939]

[6]. Rancati T, Schwarz M, Allen AM, Feng F, Popovtzer A, Mittal B, et al. Radiation Dose–Volume Effects in the Larynx and Pharynx. International Journal of Radiation Oncology*Biology*Physics 2010;76:S64–9. 10.1016/j.ijrobp.2009.03.079.

[7]. Deasy JO, Moiseenko V, Marks L, Chao KSC, Nam J, Eisbruch A. Radiotherapy Dose-Volume Effects on Salivary Gland Function. Int J Radiat Oncol Biol Phys 2010;76:S58. 10.1016/J.IJROBP.2009.06.090. [PubMed: 20171519]

[8]. Eisbruch A, Kim HM, Terrell JE, Marsh LH, Dawson LA, Ship JA. Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. International Journal of Radiation Oncology*Biology*Physics 2001;50:695–704. 10.1016/S0360-3016(01)01512-7.

[9]. Beetz I, Schilstra C, Van Der Schaaf A, Van Den Heuvel ER, Doornaert P, Van Luijk P, et al. NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: The role of dosimetric and clinical factors. Radiother Oncol 2012;105:101–6. 10.1016/j.radonc.2012.03.004. [PubMed: 22516776]

[10]. Kierkels RGJ, Korevaar EW, Steenbakkers RJHM, Janssen T, van't Veld AA, Langendijk JA, et al. 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. Radiotherapy and Oncology 2014;112:430–6. 10.1016/j.radonc.2014.08.020. [PubMed: 25220369]

[11]. Brodin NP, Tomé WA. Revisiting the dose constraints for head and neck OARs in the current era of IMRT. Oral Oncol 2018;86:8–18. 10.1016/j.oraloncology.2018.08.018. [PubMed: 30409324]

[12]. 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;107:267–73. 10.1016/j.radonc.2013.05.007. [PubMed: 23759662]

[13]. Langendijk JA, Hoebers FJP, de Jong MA, Doornaert P, Terhaard CHJ, Steenbakkers RJHM, et al. National Protocol for Model-Based Selection for Proton Therapy in Head and Neck Cancer. Int J Part Ther 2021;8:354–65. 10.14338/IJPT-20-00089.1. [PubMed: 34285961]

[14]. Christianen MEMC, Schilstra C, Beetz I, Muijs CT, Chouvalova O, Burlage FR, Doornaert P, Koken PW, Leemans CR, Rinkel RNPM, De Bruijn MJ, De Bock GH, Roodenburg JLN, Van Der Laan BFAM, Slotman BJ, Verdonck-De Leeuw IM, Bijl HP, Langendijk JA. Predictive modelling for swallowing dysfunction after primary (chemo)radiation: Results of a prospective observational study. Radiotherapy and Oncology 2012;105:107–14. 10.1016/j.radonc.2011.08.009. [PubMed: 21907437]

[15]. Sapir E, Tao Y, Feng F, Samuels S, El Naqa I, Murdoch-Kinch CA, et al. Predictors of Dysgeusia in Patients With Oropharyngeal Cancer Treated With Chemotherapy and Intensity Modulated Radiation Therapy. International Journal of Radiation Oncology*Biology*Physics 2016;96:354–61. 10.1016/j.ijrobp.2016.05.011.

[16]. Rosenthal DI, Chambers MS, Fuller CD, Rebueno NCS, Garcia J, Kies MS, et al. Beam Path Toxicities to Non-Target Structures During Intensity-Modulated Radiation Therapy for Head and Neck Cancer. International Journal of Radiation Oncology*Biology*Physics 2008;72:747–55. 10.1016/j.ijrobp.2008.01.012.

[17]. Van den Bosch L, van der Schaaf A, van der Laan HP, Hoebers FJP, Wijers OB, van den Hoek JGM, et al. Comprehensive toxicity risk profiling in radiation therapy for head and neck cancer: A new concept for individually optimised treatment. Radiother Oncol 2021;157:147–54. 10.1016/j.radonc.2021.01.024. [PubMed: 33545258]

[18]. Van den Bosch L, Schuit E, van der Laan HP, Reitsma JB, Moons KGM, Steenbakkers RJHM, et al. Key challenges in normal tissue complication probability model development and validation: towards a comprehensive strategy. Radiother Oncol 2020;148:151–6. 10.1016/j.radonc.2020.04.012. [PubMed: 32388149]

[19]. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med 2015;162:W1–73. 10.7326/M14-0698. [PubMed: 25560730]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[20]. Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, et al. Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): An Introduction to the Scientific Issues. International Journal of Radiation Oncology*Biology*Physics 2010;76:S3–9. 10.1016/j.ijrobp.2009.09.040.

[21]. Mohamed ASR, Ruangskul M-N, Awan MJ, Baron CA, Kalpathy-Cramer J, Castillo R, et al. Quality Assurance Assessment of Diagnostic and Radiation Therapy–Simulation CT Image Registration for Head and Neck Radiation Therapy: Anatomic Region of Interest–based Comparison of Rigid and Deformable Algorithms. Radiology 2015;274:752–63. 10.1148/radiol.14132871. [PubMed: 25380454]

[22]. Brouwer CL, Steenbakkers RJHM, Bourhis J, Budach W, Grau C, Grégoire V, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. Radiother Oncol 2015;117:83–90. 10.1016/j.radonc.2015.07.041. [PubMed: 26277855]

[23]. van Dijk LV, Van den Bosch L, Aljabar P, Peressutti D, Both S, Roel JHM, et al. Improving automatic delineation for head and neck organs at risk by Deep Learning Contouring. Radiotherapy and Oncology 2020;142:115–23. https://doi.sorg/10.1016/j.radonc.2019.09.022. [PubMed: 31653573]

[24]. List MA, Ritter-Sterr C, Lansky SB. A performance status scale for head and neck cancer patients. Cancer 1990;66:564–9. 10.1002/1097-0142(19900801)66:3<564::AID-CNCR2820660326>3.0.CO;2-D. [PubMed: 2364368]

[25]. Rosenthal DI, Mendoza TR, Chambers MS, Asper JA, Gning I, Kies MS, et al. Measuring head and neck cancer symptom burden: The development and validation of the M. D. Anderson symptom inventory, head and neck module. Head Neck 2007;29:923–31. 10.1002/hed.20602. [PubMed: 17358040]

[26]. Bjordal K, de Graeff A, Fayers PM, Hammerlid E, van Pottelsberghe C, Curran D, et al. A 12 country field study of the EORTC QLQ-C30 (version 3.0) and the head and neck cancer specific module (EORTC QLQ-H&N35) in head and neck patients. Eur J Cancer 2000;36:1796–807. 10.1016/S0959-8049(00)00186-6. [PubMed: 10974628]

[27]. US Department of Health and Human Services National Institutes of Health National Cancer Institute, Common Terminology Criteria for Adverse Events (CTCAE) Version 5, (2017). https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/ctcae_v5_quick_reference_5x7.pdf (accessed October 27, 2022).

[28]. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol 2016;74:167–76. 10.1016/j.jclinepi.2015.12.005. [PubMed: 26772608]

[29]. R.C. Team, R: A language and environment for statistical computing, (2022). https://www.r-project.org/.

[30]. van der Laan HP, van der Schaaf A, Van den Bosch L, Korevaar EW, Steenbakkers RJHM, Both S, et al. Quality of life and toxicity guided treatment plan optimisation for head and neck cancer. Radiother Oncol 2021;162:85–90. 10.1016/j.radonc.2021.06.035. [PubMed: 34237344]

[31]. van der Laan HP, Gawryszuk A, van der Schaaf A, Langendijk JA. Risk reduction of radiation-induced aspiration by sparing specific aspiration-related-organs at risk; an in silico feasibility study. Radiother Oncol 2025;203:110698. 10.1016/j.radonc.2024.110698. [PubMed: 39716592]

[32]. van Bruggen IG, Huiskes M, de Vette SPM, Holmström M, Langendijk JA, Both S, et al. Automated Robust Planning for IMPT in Oropharyngeal Cancer Patients Using Machine Learning. International Journal of Radiation Oncology*Biology*Physics 2023;115:1283–90. 10.1016/j.ijrobp.2022.12.004.

[33]. Hansen CR, Jensen K, Smulders B, Holm AIS, Samsøe E, Nielsen MS, Sibolt P, Skyt P, Elstrøm UV, Nielsen CP, Johansen J, Zukauskaite R, Eriksen JG, Farhadi M, Andersen M, Andersen E, Overgaard J, Grau C, Friborg J. Evaluation of decentralised model-based selection of head and neck cancer patients for a proton treatment study. DAHANCA 35. Radiotherapy and Oncology 2024;190:109812. 10.1016/j.radonc.2023.109812. [PubMed: 37479061]

[34]. Gan Y, Langendijk JA, van der Schaaf A, van den Bosch L, Oldehinkel E, Lin Z, et al. An efficient strategy to select head and neck cancer patients for adaptive radiotherapy. Radiother Oncol 2023;186:109763. 10.1016/j.radonc.2023.109763. [PubMed: 37353058]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[35]. Walser MA, Bachmann N, Kluckert J, Köthe A, Tully C, Leiser D, et al. Clinical outcome after pencil beam scanning proton therapy and dysphagia/xerostomia NTCP calculations of proton and photon radiotherapy delivered to patients with cancer of the major salivary glands. Br J Radiol 2023;96. 10.1259/bjr.20220672.

[36]. Kaplan LP, Holm AIS, Eriksen JG, Heijmen BJM, Korreman SS, Rossi L. Patient anatomy-specific trade-offs between sub-clinical disease coverage and normal tissue dose reduction in head-and-neck cancer. Radiother Oncol 2023;182:109526. 10.1016/j.radonc.2023.109526. [PubMed: 36764458]

[37]. Tambas M, van der Laan HP, Steenbakkers RJHM, Doyen J, Timmermann B, Orlandi E, et al. Current practice in proton therapy delivery in adult cancer patients across Europe. Radiother Oncol 2022;167:7–13. 10.1016/j.radonc.2021.12.004. [PubMed: 34902370]

[38]. Gawryszuk A, Bijl HP, Holwerda M, Halmos GB, Wedman J, Witjes MJH, et al. Functional Swallowing Units (FSUs) as organs-at-risk for radiotherapy. PART 1: Physiology and anatomy. Radiotherapy and Oncology 2019;130:62–7. 10.1016/j.radonc.2018.10.028. [PubMed: 30420235]

[39]. Barbon CEA, Peterson CB, Moreno AC, Lai SY, Reddy JP, Sahli A, et al. Adhering to Eat and Exercise Status During Radiotherapy for Oropharyngeal Cancer for Prevention and Mitigation of Radiotherapy-Associated Dysphagia. JAMA Otolaryngology–Head & Neck Surgery 2022;148:956. 10.1001/jamaoto.2022.2313. [PubMed: 36074459]

[40]. Ellsworth SG, van Rossum PSN, Mohan R, Lin SH, Grassberger C, Hobbs B. Declarations of Independence: How Embedded Multicollinearity Errors Affect Dosimetric and Other Complex Analyses in Radiation Oncology. International Journal of Radiation Oncology*Biology*Physics 2023;117:1054–62. 10.1016/j.ijrobp.2023.06.015.

[41]. European Medicines Agency, RESTART; Autologous Transplantation of Adult Salivary Gland Stem Cells to Restore Submandibular Gland Function after Radiotherapy [Internet], EU Clinical Trials Register (2024). https://euclinicaltrials.eu/search-for-clinical-trials/?lang=en&EUCT=2024-512968-57-00 (accessed February 7, 2025).

[42]. Boesveldt S, Parma V. The importance of the olfactory system in human well-being, through nutrition and social behavior. Cell Tissue Res 2021;383:559–67. 10.1007/s00441-020-03367-7. [PubMed: 33433688]

[43]. Pedersen A, Bardow A, Jensen SB, Nauntofte B. Saliva and gastrointestinal functions of taste, mastication, swallowing and digestion. Oral Dis 2002;8:117–29. 10.1034/j.1601-0825.2002.02851.x. [PubMed: 12108756]

[44]. Gunn L, Gilbert J, Nenclares P, Soliman H, Newbold K, Bhide S, et al. Taste dysfunction following radiotherapy to the head and neck: A systematic review. Radiother Oncol 2021;157:130–40. 10.1016/j.radonc.2021.01.021. [PubMed: 33545253]

[45]. Chu H, de Vette SPM, Neh H, Sijtsema NM, Steenbakkers RJHM, Moreno A, et al. Three-Dimensional Deep Learning Normal Tissue Complication Probability Model to Predict Late Xerostomia in Patients With Head and Neck Cancer. International Journal of Radiation Oncology*Biology*Physics 2024. 10.1016/j.ijrobp.2024.07.2334.

[46]. Salama V, Youssef S, Xu T, Wahid KA, Chen J, Rigert J, et al. Temporal characterization of acute pain and toxicity kinetics during radiation therapy for head and neck cancer. A retrospective study. Oral Oncology Reports 2023;7:s100092. 10.1016/j.oor.2023.100092.

[47]. Rodriguez AM, Komar A, Ringash J, Chan C, Davis AM, Jones J, et al. A scoping review of rehabilitation interventions for survivors of head and neck cancer. Disabil Rehabil 2019;41:2093–107. 10.1080/09638288.2018.1459880. [PubMed: 29976091]

[48]. Chow JCH, Cheung K-M, Cheung GTC, Tam AHP, Lui JCF, Lee FKH, et al. Dose-volume predictors of post-radiation primary hypothyroidism in head and neck cancer: A systematic review. Clin Transl Radiat Oncol 2022;33:83–92. 10.1016/j.ctro.2022.01.001. [PubMed: 35128087]

[49]. Kalendralis P, Sloep M, Moni George N, Snel J, Veugen J, Hoebers F, et al. Independent validation of a dysphagia dose response model for the selection of head and neck cancer patients to proton therapy. Phys Imaging Radiat Oncol 2022;24:47–52. 10.1016/j.phro.2022.09.005. [PubMed: 36158240]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[50]. Christianen MEMC, van der Schaaf A, van der Laan HP, Verdonck-de Leeuw IM, Doornaert P, Chouvalova O, et al. Swallowing sparing intensity modulated radiotherapy (SW-IMRT) in head and neck cancer: Clinical validation according to the model-based approach. Radiotherapy and Oncology 2016;118:298–303. 10.1016/j.radonc.2015.11.009. [PubMed: 26700602]

[51]. Hansen CR, Friborg J, Jensen K, Samsøe E, Johnsen L, Zukauskaite R, et al. NTCP model validation method for DAHANCA patient selection of protons versus photons in head and neck cancer radiotherapy. Acta Oncol (Madr) 2019;58:1410–5. 10.1080/0284186X.2019.1654129.

[52]. Huynh T-T-M, Dale E, Falk RS, Hellebust TP, Astrup GL, Malinen E, et al. Radiation-induced long-term dysphagia in survivors of head and neck cancer and association with dose-volume parameters. Radiother Oncol 2024;190:110044. 10.1016/j.radonc.2023.110044. [PubMed: 38061420]

[53]. Blanchard P, Wong AJ, Gunn GB, Garden AS, Mohamed ASR, Rosenthal DI, et al. Toward a model-based patient selection strategy for proton therapy: External validation of photon-derived normal tissue complication probability models in a head and neck proton therapy cohort. Radiother Oncol 2016;121:381–6. 10.1016/j.radonc.2016.08.022. [PubMed: 27641784]

[54]. Kamal M, Rosenthal DI, Volpe S, Goepfert RP, Garden AS, Hutcheson KA, et al. Patient reported dry mouth: Instrument comparison and model performance for correlation with quality of life in head and neck cancer survivors. Radiotherapy and Oncology 2018;126:75–80. 10.1016/j.radonc.2017.10.037. [PubMed: 29229507]

[55]. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med 2016; 35:214–26. 10.1002/sim.6787. [PubMed: 26553135]

[56]. Tambas M Patient Selection for Proton Therapy in Head and Neck Cancer. University of Groningen; 2023.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

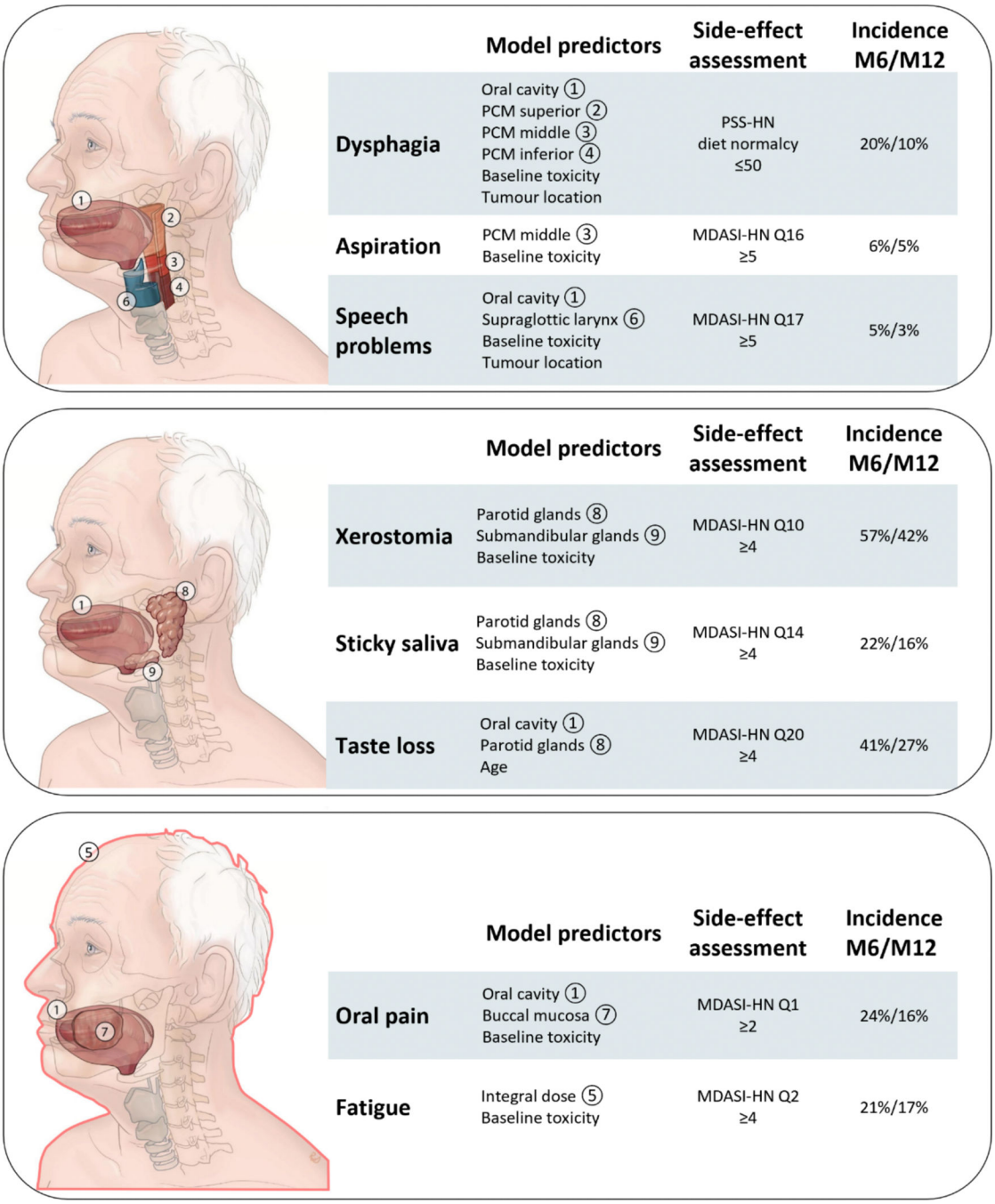

Fig. 1. CITOR models: selected predictors per model, side-effect assessment and incidence at 6 and 12 months after radiotherapy.

Adapted from Van den Bosch et al. [17]. Organs at risk identified as predictors: 1 = oral cavity; 2 = superior PCM; 3 = middle PCM; 4 = inferior PCM; 5 = integral dose; 6 = supraglottic larynx; 7 = buccal mucosa*; 8 = parotid gland*; 9 = submandibular gland*. *For paired structures only one side is depicted. Abbreviations: PCM = pharyngeal constrictor muscle; M6 = 6 months after radiotherapy; M12 = 12 months after radiotherapy; PSS-HN=Performance Status Scale for Head and Neck cancer; MDASI-HN Q = MD Anderson Symptom Inventory Head and Neck Question.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

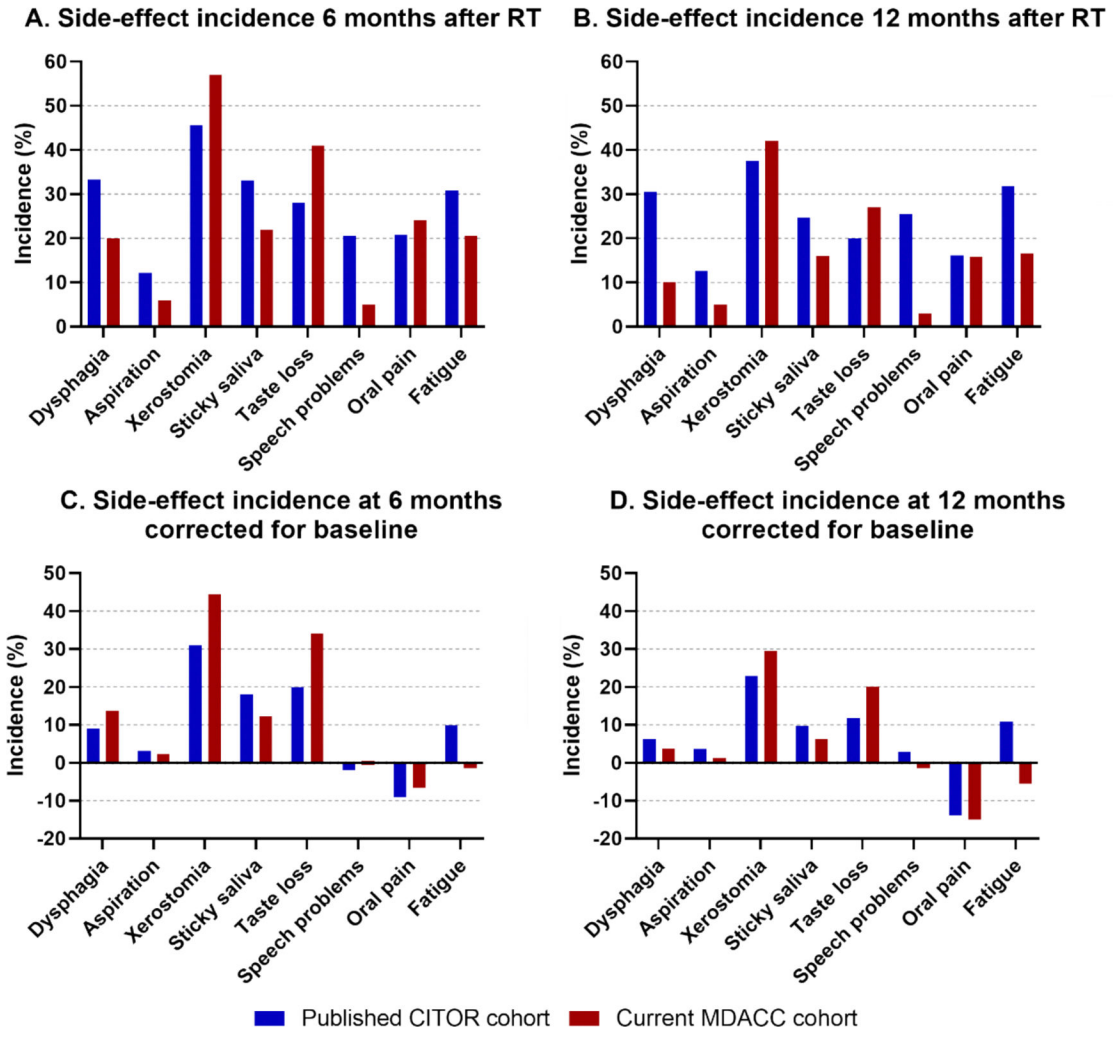

Fig. 2.

Side-effect incidences. Side-effect incidences at 6 and 12 months after RT (A-B) and after subtracting incidence at baseline (C-D) of the MDACC cohort and the previously published CITOR development cohort [17]. Abbreviations: CITOR = comprehensive individual toxicity risk; MDACC = MD Anderson Cancer Center; RT = radiation therapy.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

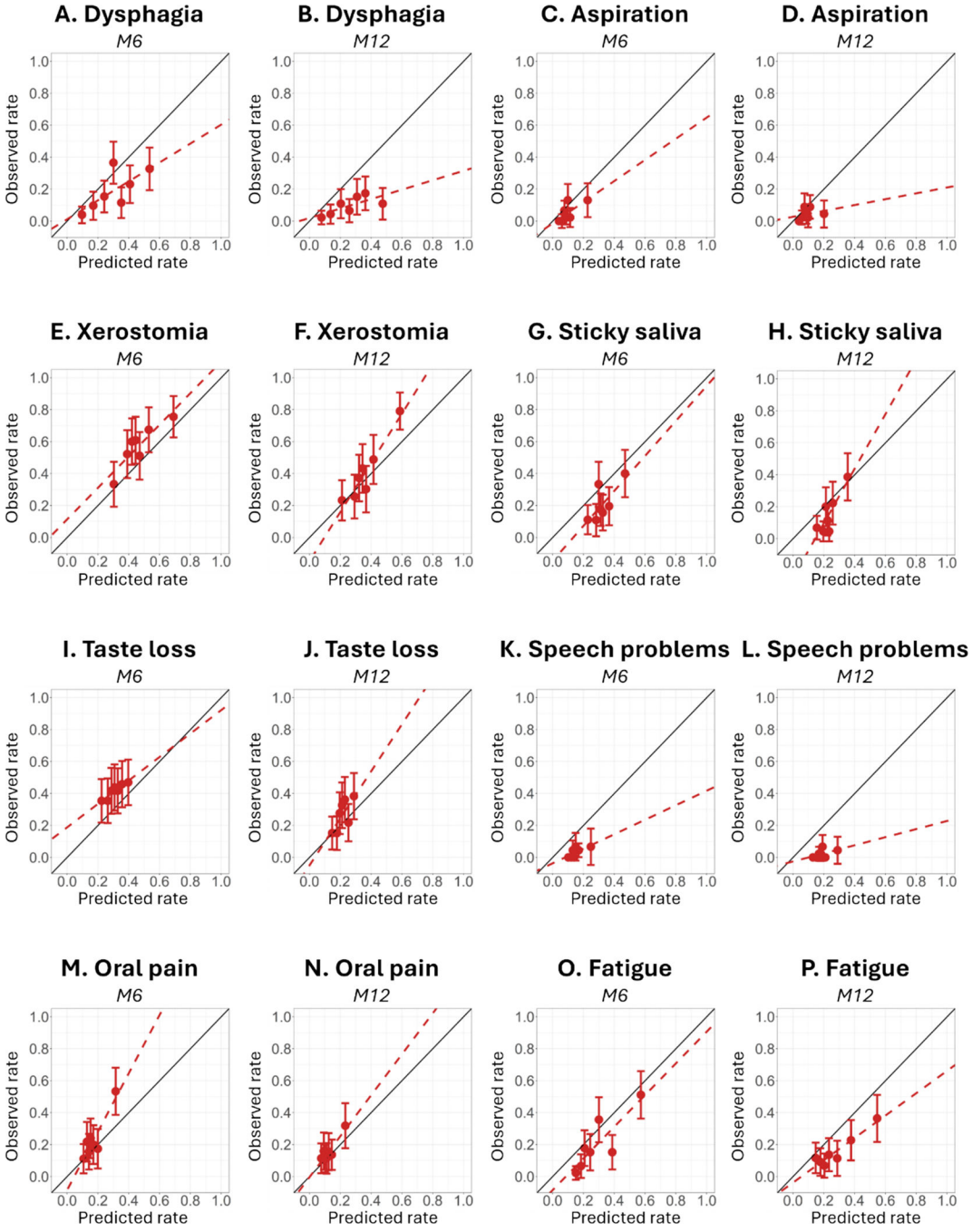

Fig. 3.

Calibration plots. Calibration plots of models for all side-effects at 6 and 12 months after radiation therapy. *Abbreviations:* M6 = 6 months after radiation therapy; M12 = 12 months after radiation therapy.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

MDACC cohort compared to CITOR development cohort.

	Current MDACC cohort (n = 407)	Published CITOR cohort[17] (n = 750)	P value
Clinical characteristics			
Age in years, mean (sd)	60 (9)	63 (10)	<0.001 *‡
Sex, No. (%)			<0.001 *†
Male	366 (89.9)	560 (74.7)	
Female	41 (10.1)	190 (25.3)	
Tumor site, No. (%)			<0.001 *//
Oral cavity	0 (0.0)	44 (5.9)	
Oropharynx	370 (90.9)	271 (36.1)	
Nasopharynx	0 (0.0)	30 (4.0)	
Hypopharynx	0 (0.0)	71 (9.5)	
Larynx	0 (0.0)	334 (44.5)	
Unknown primary	37 (9.1)	0 (0.0)	
HPV status (only in oropharyngeal cancer patients), No. (%)			<0.001 *†
Positive	275 (74.3)	98 (36.2)	
Negative	17 (4.6)	151 (55.7)	
Not tested	78 (21.1)	22 (8.1)	
Tumor stage, No. (%)			<0.001 *†
Tis-T2	295 (72.5)	363 (48.4)	
T3-T4	111 (27.3)	387 (51.6)	
Nodal stage, No. (%)			<0.001 *†
N0	31 (7.6)	333 (44.4)	
N1	152 (37.4)	64 (8.5)	
N2	216 (53.2)	330 (44.0)	
N3	7 (1.7)	23 (3.1)	
Treatment technique, No. (%)			<0.001 *//
3D-CRT	0 (0.0)	86 (11.5)	
IMRT	68 (16.7)	546 (72.8)	
VMAT	232 (57.1)	106 (14.1)	
IMRT + VMAT	32 (7.9)	12 (1.6)	
IMPT	74 (18.2)	0 (0)	
Treatment modality, No. (%)			<0.001 *†
RT alone	93 (22.9)	443 (59.1)	
RT with systemic therapy	307 (75.4)	307 (40.9)	
Mean physical dose to OAR in Gy, median (IQR)			
Ipsilateral buccal mucosa	44.8 (33.8;53.6)	38.5 (12.3;54.6)	<0.001 *≠
Contralateral buccal mucosa	27.2 (17.6;35.1)	30.9 (8.9;42.0)	0.04 *≠

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

	Current MDACC cohort (n = 407)	Published CITOR cohort[17] (n = 750)	P value
Oral cavity	46.5 (40.6;52.4)	43.9 (22.3;55.6)	0.001 *≠
Ipsilateral parotid gland	31.7 (26.3;37.3)	30.7 (18.3;42.1)	0.09 ≠
Contralateral parotid gland	17.8 (13.6;21.5)	24.3 (12.7;31.5)	<0.001 *≠
Inferior PCM	34.8 (25.0;43.7)	57.6 (45.8;66.3)	<0.001 *≠
Middle PCM	55.2 (43.3;61.9)	55.7 (41.2;64.3)	0.77 ≠
Superior PCM	61.8 (56.1;66.0)	52.3 (29.3;62.6)	<0.001 *≠
Ipsilateral submandibular gland	70.4 (67.6;71.6)	63.7 (47.6;68.3)	<0.001 *≠
Contralateral submandibular gland	51.0 (37.8;60.9)	53.0 (42.5;62.0)	0.48 ≠
Supraglottic area	50.5 (38.8;59.7)	57.9 (46.1;66.4)	<0.001 *≠
Integral dose in Gy, median (IQR)	$2.0*10^5$ ($1.6*10^5$;$2.3*10^5$)	$1.6*10^5$ ($1.1*10^5$;$2.0*10^5$)	<0.001 *≠
Volume external contour in cc	$1.7*10^4$ ($1.5*10^4$;$1.9*10^4$)	$1.2*10^4$ ($1.0*10^4$;$1.4*10^4$)	<0.001 *≠

Abbreviations. 3D-CRT = 3-dimensional conformal radiation therapy; CITOR = comprehensive individual toxicity risk; Gy = gray; HPV = human papilloma virus; IMPT = intensity modulated proton therapy; IMRT = intensity modulated radiation therapy; IQR = interquartile range; MDACC = MD Anderson Cancer Center; OAR = organ at risk; PCM = pharyngeal constrictor muscle; sd = standard deviation; VMAT = volumetric modulated arc therapy.

*Statistically significant ($\alpha \leq 0.05$) using

≠One-way analysis of variance test,

†Fisher's exact test,

//χ^2 test, and

≠Kruskal-Wallis test.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Definition side-effects in MDACC cohort.

	Overall correlation		Appropriate cut-off value in MDACC cohort				
	Spearman's ρ [95 % CI]	**P value**	**Cut-off value**	**Youden's J Index**	**Sensitivity**	**Specificity**	**Cohen's *k* [95 % CI]**
Dysphagia	0.85 [0.83;0.88]	<0.001	≤50	0.92	0.94	0.99	0.93 [0.90;0.96]
Aspiration	0.61 [0.59;0.63]	<0.001	≥5	0.72	0.75	0.97	0.50 [0.29;0.70]
Xerostomia	0.83 [0.81;0.85]	<0.001	≥4	0.69	0.85	0.85	0.65 [0.58;0.73]
Sticky saliva	0.64 [0.62;0.66]	<0.001	≥4	0.59	0.73	0.86	0.54 [0.44;0.63]
Taste loss	0.84 [0.81;0.86]	<0.001	≥4	0.72	0.78	0.94	0.70 [0.61;0.78]
Speech problems	0.70 [0.68;0.72]	<0.001	≥5	0.71	0.79	0.92	0.65 [0.56;0.75]
Oral pain	0.52 [0.50;0.54]	<0.001	≥2	0.62	0.93	0.69	0.43 [0.35;0.50]
Fatigue	0.78 [0.76;0.80]	<0.001	≥4	0.67	0.96	0.71	0.39 [0.30;0.49]

Abbreviations. CI = confidence interval; MDACC = MD Anderson Cancer Center; NTCP = normal tissue complication probability.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Oral Oncol. Author manuscript; available in PMC 2025 April 01.

Table 3

Performance of NTCP models in MDACC cohort.

					R^2	Calibration	
		No. patients	**No. events (incidence [%])**	**AUC [95 % CI]**		**Intercept [95 % CI]**	**Slope [95 % CI]**
Swallowing domain							
Dysphagia	M06	372	73 (20)	0.66 [0.60;0.73]	0.108	−0.75 [−1.10;−0.40]	0.89 [0.51;1.24]
	M12	329	32 (10)	0.64 [0.55;0.74]	0.046	−1.60 [−2.15;−1.05]	0.64 [0.15;1.13]
Aspiration	M06	327	19 (6)	0.72 [0.62;0.83]	0.091	−0.55 [−1.81;0.70]	1.08 [0.45;1.71]
	M12	322	16 (5)	0.62 [0.49;0.76]	0.038	−1.32 [−2.84;0.21]	0.75 [0.04;1.45]
Salivary domain							
Xerostomia	M06	326	186 (57)	0.63 [0.56;0.69]	0.080	0.43 [0.19;0.67]	0.99 [0.52;1.46]
	M12	310	129 (42)	0.69 [0.63;0.75]	0.164	0.47 [0.10;0.84]	1.47 [0.96;1.99]
Sticky saliva	M06	326	72 (22)	0.63 [0.56;0.71]	0.084	−0.21 [−0.74;0.33]	1.57 [0.83;2.30]
	M12	318	52 (16)	0.70 [0.62;0.78]	0.126	0.64 [−0.28;1.56]	2.04 [1.22;2.87]
Taste loss	M06	345	143 (41)	0.55 [0.49;0.61]	0.015	0.23 [−0.40;0.86]	0.72 [−0.03;1.46]
	M12	333	91 (27)	0.61 [0.54;0.67]	0.043	0.77 [−0.37;1.92]	1.37 [0.48;2.26]
Speech domain							
Speech problems	M06	323	15 (5)	0.80 [0.68;0.93]	0.284	0.64 [−0.63;1.92]	2.49 [1.57;3.41]
	M12	321	9 (3)	0.75 [0.59;0.90]	0.134	−1.15 [−2.46;0.15]	1.89 [0.82;2.96]
Pain domain							
Oral pain	M06	326	79 (24)	0.66 [0.59;0.73]	0.132	1.11 [0.26;1.96]	1.48 [0.93;2.03]
	M12	317	50 (16)	0.60 [0.51;0.69]	0.043	0.02 [−1.13;1.18]	0.89 [0.29;1.48]
General domain							
Fatigue	M06	326	67 (21)	0.73 [0.67;0.80]	0.164	−0.53 [−0.90;−0.15]	1.08 [0.71;1.45]
	M12	315	52 (17)	0.68 [0.59;0.76]	0.098	−0.91 [−1.32;−0.50]	0.86 [0.47;1.25]

Abbreviations. AUC = area under the curve; CI = confidence interval; MDACC = MD Anderson Cancer Center; M06 = 6 months after radiation therapy; M12 = 12 months after radiation therapy; NTCP = normal tissue complication probability; R^2 = explained variance.