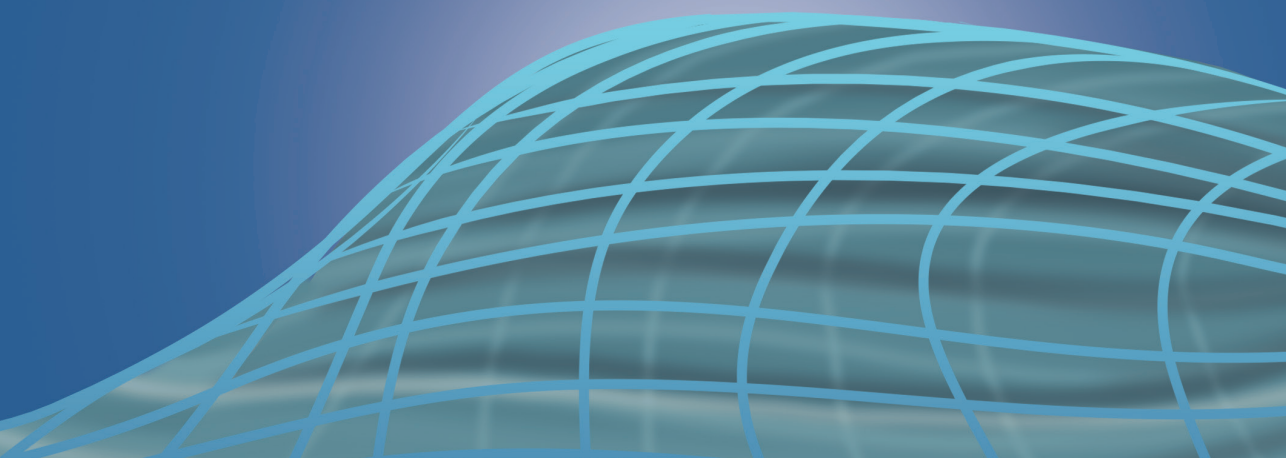
CHAPTER 3

Direct use of multivariable normal tissue complication probability models in treatment plan optimization for individualized head and neck cancer radiotherapy produces clinically acceptable treatment plans

Kierkels, Roel G J
Korevaar, Erik W
Steenbakkers, Roel J H M
Janssen, Tomas
Van't Veld, Aart A
Langendijk, Johannes A
Schilstra, Cornelis
van der Schaaf, Arjen

Radiotherapy and oncology 112 (2014) 430-436



Abstract

Background and purpose

Recently, clinically validated multivariable normal tissue complication probability models (NTCP) for head and neck cancer (HNC) patients have become available. We test the feasibility of using multivariable NTCP-models directly in the optimiser for inverse treatment planning of radiotherapy to improve the dose distributions and corresponding NTCP-estimates in HNC patients.

Material and Methods

For 10 HNC cases, intensity-modulated radiotherapy plans were optimised either using objective functions based on the 'generalised equivalent uniform dose' (OF_{gEUD}) or based on multivariable NTCP-models (OF_{NTCP}). NTCP-models for patient-rated xerostomia, physician-rated RTOG grade 2-4 dysphagia, and various patient-rated aspects of swallowing dysfunction were incorporated. The NTCP-models included dose-volume parameters as well as clinical factors contributing to a personalised optimisation process. Both optimisation techniques were compared by means of 'pseudo Pareto fronts' (target dose conformity vs. the sum of the NTCPs).

Results

Both optimisation techniques resulted in clinically realistic treatment plans with only small differences. For nine patients the sum-NTCP was lower for the OF_{NTCP} optimised plans (on average 5.7% (95%CI 1.7–9.9%, $p<0.006$)). Furthermore, the OF_{NTCP} provided the advantages of fewer unknown optimisation parameters and an intrinsic mechanism of individualisation.

Conclusions

Treatment plan optimisation using multivariable NTCP-models directly in the OF is feasible as has been demonstrated for HNC radiotherapy.

3.1 Introduction

Radiotherapy plays a pivotal role in the curative treatment of head and neck cancer (HNC) patients. The general goal is to achieve locoregional tumour control and to preserve function of healthy organs. Quality of life after HNC radiotherapy is significantly affected by the levels of xerostomia and dysphagia [1-3]. The introduction of intensity-modulated radiotherapy (IMRT) for HNC resulted in decreased incidences of radiation-induced side effects [4-6]. To achieve this reduction, IMRT plans are optimised ensuring sufficient dose coverage to the planning target volume (PTV) while minimizing the doses to surrounding healthy tissues.

Generally, dose distributions are optimised by minimizing an objective function (OF) based on dose-volume parameters. The clinical impact of the physical dose distribution is, however, not described explicitly with this type of OF. In addition, the relationship between physical dose and the response of tumour and healthy tissue is non-linear [7]. Plan optimisation and evaluation are therefore increasingly based on biological indices accounting for dose response relationships. Guidelines on the use of biologically related models for treatment planning have recently been reported by Allen Li *et al.* [8].

Dose response relationships for tumour control probability and normal tissue complication probability (NTCP) are commonly described as sigmoidal functions of dose. A commonly used class of NTCP-models utilises the Lyman-Kutcher-Burman (LKB) model, considering heterogeneous dose distributions by use of the 'generalised equivalent uniform dose' (gEUD) [9]. The gEUD uses the physical dose distribution and a tissue-dependent parameter a to describe tumour or healthy tissue properties [10]. Currently, the gEUD formalism is the most frequently used biologically motivated OF in commercial treatment planning systems (TPS)[11].

Several studies demonstrated successful application of gEUD-based OFs for normal tissue sparing [12-18]. However, application of gEUD-based OFs to the PTVs can result in undesirable hotspots when no additional gEUD-based objective with a positive volume effect parameter to control the overdosage is added [17]. These hotspots can be prevented by so-called hybrid OFs, including dose-based objectives for the targets and gEUD-based objectives for the organs at risk (OAR) [13,14].

The gEUD can directly be converted into an NTCP using the LKB-model. However, the use of more sophisticated multivariable NTCP-models would be more clinically meaningful than the use of the gEUD alone. Recently, multivariable logistic NTCP-models were developed for xerostomia and swallowing dysfunction, including multiple dose-volume

variables and clinical factors [19-21]. We expect that these multivariable NTCP-models will be increasingly used in routine clinical practice, e.g. to select patients who may benefit most from new emerging treatment modalities, such as proton therapy [22].

In a recently published in silico comparative planning study, the application of swallowing sparing IMRT reduced NTCP-estimates for dysphagia compared to standard IMRT [23]. Plans were optimised using a hybrid gEUD-based OF (private communication with C.S.). The resulting plans were evaluated with multivariable NTCP-models for swallowing dysfunction. It is conceivable that direct application of these NTCP-models in the optimisation process would further improve plan quality. This could be due to the lower number of parameters in the OF for NTCP-based optimisations than for gEUD-based optimisations. Therefore, a treatment plan that is optimal with respect to gEUD is not necessarily optimal with respect to NTCP. Furthermore, multivariable NTCP-models depend on a predefined combination of doses in multiple OARs and on additional prognostic clinical factors, while a gEUD-based OF only includes dose information for separate OARs without predefined weights and no additional clinical factors. Also, NTCP-values are proportional to clinically relevant responses, while dosimetric variables are not. This facilitates optimal plan selection for each individual patient.

The aim of this study was to demonstrate radiotherapy treatment planning optimisation with direct use of multivariable NTCP-models for HNC patients.

3.2 Materials and Methods

3.2.1 Patients

The study population of this study composed of 9 males and 1 female (median age 59; range: 54–82 years), diagnosed with stage II–IV squamous cell HNC. All patients were consecutively selected from a database of HNC patients included in a prospective data registration program. The patients selected had tumours that originated in the oropharynx, larynx and hypopharynx. All patients were previously treated with curatively intended IMRT either alone or in combination with concomitant chemotherapy or cetuximab. All patients were previously included in recent published studies [20,21].

3.2.2 Prescription and delineation

For each patient, a simultaneous integrated boost technique was planned comprising two dose level prescriptions: 70 Gy to the PTV (PTV_{boost}, in 2 Gy per fraction, 5 fractions per week in 7 weeks) and 54.25 Gy ($PTV_{prophylactic}$, in 1.55 Gy per fraction) to the prophylactic lymph node regions in both sides of the neck. Both PTVs consisted of a clinical target volume

with 5 mm margin for internal and setup uncertainties. OARs involved in radiation-induced salivary and swallowing dysfunction were delineated according to previously published guidelines [24,25]. Additionally, the spinal cord, brain and a ring of 1 cm around $PTV_{prophylactic}$ were contoured for plan optimisation and evaluation.

3.2.3 Objective functions and NTCP-models

IMRT plans were optimised with two different hybrid OFs (Supplementary material). Both OFs included physical dose objectives for the PTVs. For the OARs, the first OF (OF_{gEUD}) utilised gEUD objectives. For the second OF (OF_{NTCP}) multivariable NTCP-models were assigned to the OARs. NTCP-models were included for patient-reported moderate to severe xerostomia, physician-rated grade II-IV dysphagia, and patient-reported swallowing problems related to consumption of solid food, soft food or liquids and choking[20,21]. The corresponding variables and regression coefficients are listed in table S.31 (Supplementary material).

3.2.4 Treatment planning

All treatment plans were created in the Pinnacle^3 TPS (research version (v9.1), Philips Healthcare, Andover, MA) and consisted of seven equidistant 6 MV beams. Plans were optimised with OF_{gEUD} and OF_{NTCP}. For both OFs, identical target structures and OARs were used. Target requirements for the plan evaluation phase included $V_{66.5Gy} \geq 98\%$ and $V_{74.9Gy} \leq 2\%$ for PTV_{boost} and $V_{51.5Gy} \geq 98\%$ for $PTV_{prophylactic}$. The maximum dose (D0%) allowed to the spinal cord and the brain was 54 and 60 Gy, respectively. The additional OARs in the OFs were related to xerostomia and swallowing dysfunction (table 3.S1).

For OF_{gEUD}-plans, the required gEUD thresholds were chosen based on an initial optimisation run (such that the evaluation of the OF produced a value around 0.02, based on clinical experience), using low objective weights, i.e. 1-5, and a gEUD parameter of a=1 (indicating the mean dose). The parameter a was chosen to be 1 to agree with the OARs mean dose parameters in the multivariable NTCP-models. A more detailed characterisation of the dose distribution ($a \neq 1$) would not improve the gEUD-based results, because these were finally evaluated using NTCP-models that also incorporated mean dose parameters. For OF_{NTCP}-plans, the convexified multivariable NTCP-models [h(NTCP), Supplementary material] and their derivatives (for steepest-descent optimisation) were implemented as a plug-in to the Pinnacle^3 Research Interface (v1.2). Inside the inverse planning menu the h(NTCP) models were invoked via these plug-ins. All patient and treatment related clinical factors were available from our standard follow up program, and set prior to optimisation. The NTCP-based objective weights were initially set to 0.05. Relatively high objective weights, i.e. 100, were assigned to the target objectives.

In current clinical practice, the final plan quality highly depends on the set of relative weight factors set by the dosimetrist. The search for optimal weights may be time consuming, subjective, and a major source of plan variability. To deal with these problems, OF_{gEUD} and OF_{NTCP}-plans were calculated in an automated approach aiming at computing Pareto-optimal plans in which improvements of one measure can only be achieved by worsening others. Hence, a set of Pareto-optimal plans span the Pareto front (PF) [26]. Comparing PFs for the evaluation of different planning techniques, such as OF_{gEUD}- and OF_{NTCP}-based optimisations, is superior to the comparison of two single plans. In this study, plans were automatically generated using the method described by Janssen *et al*.[27].

To approximate the PFs, 200 plans were created for each optimisation technique per patient. For the first plan the planning parameters (gEUD dose thresholds and corresponding objective weights λ) were set manually, as described above. For each subsequent plan the objective parameters were varied according to adapted Gaussian distributions [27]. Moreover, the gEUD parameter a remained fixed (a=1) during the optimisation process. Each plan was optimised in 40 iterations. The first 10 iterations only consisted of fluence-map optimization [28]. In the next 30 iterations, direct aperture optimisation was performed, followed by an adaptive convolve dose computation. Contrary to the fluence-map optimisation, the direct aperture optimisation problem is non-convex, such that, in this part of the optimisation, the optimiser may be trapped in a sub-optimal local minimum. To ensure proper dose coverage, all plans were automatically re-scaled such that $D_{98\%}$=95% for PTV_{boost}.

3.2.5 Analysis

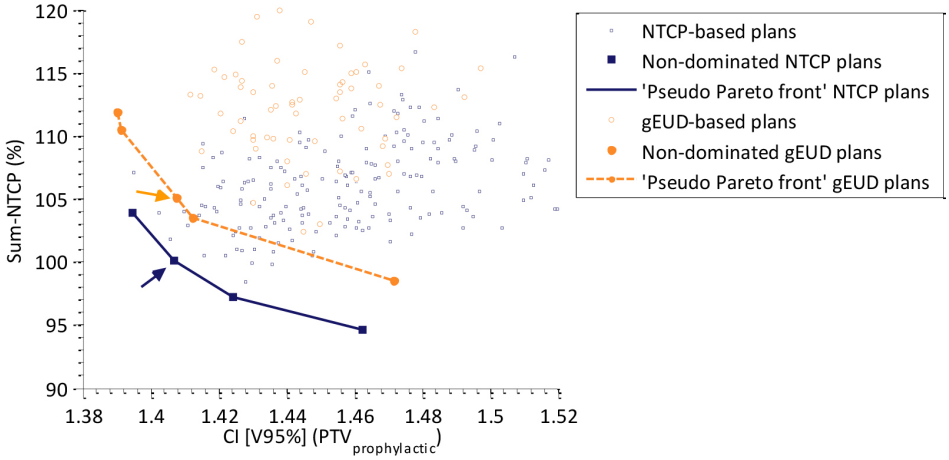
All plans and PFs were analysed and compared using in-house-developed software in MATLAB (version 7.14, Mathworks, Natick, MA). Plans were compared with respect to the conformity index ($CI_{V95\%}$), homogeneity index (HI), monitor units (MUs), dosimetric parameters and NTCPs. The $CI_{V95\%}$ was defined as the ratio between the total volume enclosed by the 95% isodose and the $V_{95\%}$ in a PTV (i.e. the volume of the PTV that received at least 95% of the prescription dose). The HI describes the dose inhomogeneity inside a PTV and was calculated according to reference [29].

All plans were evaluated on sufficient PTV coverage ($PTV_{prophylactic}$: $D_{98\%} \geq 51.5$Gy; PTV_{boost}: $D_{2\%} \leq 74.9$Gy; for PTV_{boost} the $D_{98\%}$ was not used, because for all plans the MUs were automatically re-scaled such that $D_{98\%}$=66.5Gy) and maximum dose ($D_{0\%}$) to the spinal cord. The violating plans were deleted and from the remaining plans, so-called pseudo PFs (pPFs) were constructed, projecting the $CI_{V95\%}[PTV_{prophylactic}]$ against the sum of the individual NTCP-values. The sum-NTCP was calculated with equal weights assigned to the individual

NTCP-models, since more elaborate translations of complication probabilities towards quality of life are subjective and beyond the scope of this study.

For each patient, the pPFs of both optimisation techniques were compared within the range of equal $CI_{V95\%}$ data points. Within this range, the median (sum) NTCPs, and corresponding dosimetric parameters were derived and compared. Individual differences between the median NTCPs were visualized by Bland-Altman plots. Differences between plan evaluation parameters were evaluated by two-tailed Wilcoxon signed-rank tests. Due to multiple comparisons, the level of being statistically significant ($p<0.05$) was adjusted using the Bonferroni correction and set to $p<0.05/7=0.007$. Additionally, for each patient, Pareto optimal plans of both optimisation techniques were evaluated by an expert radiation oncologist in HNC (R.J.H.M.S.).

Figure 3.1. *Comparison of 'pseudo Pareto fronts' of head and neck IMRT plans optimised with gEUD-based and multivariable NTCP-model based objective functions for one example case.*



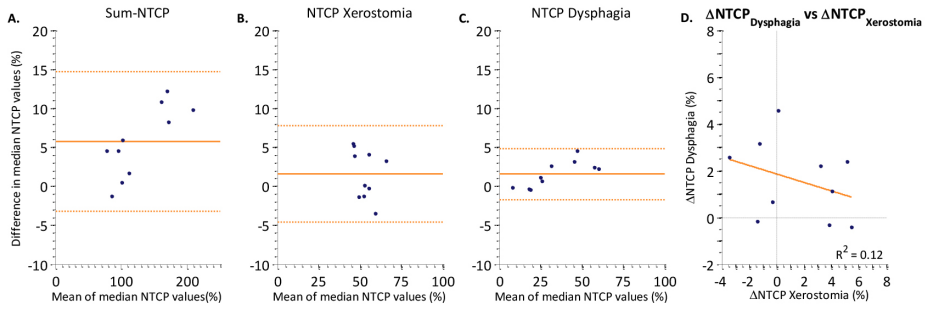
The sum of the calculated NTCP-values (see Table 3.1) is plotted against the conformity index of the 95% isodose around the $PTV_{prophylactic}$. Each data point represents one plan. The solid markers indicate the non-dominated plans. The 'pseudo Pareto fronts' for the NTCP-plans (solid line) and gEUD-plans (dashed line) are depicted as the convex hull of the non-dominated plans. The arrows point towards the plans from which the dose-volume histograms are shown in Figure 3.3.

3.3 Results

For all cases, clinically acceptable plans could be achieved (approximately 8 hours per 200 plans) using objective functions based on either gEUD objectives or multivariable NTCP-models. The number of optimal plans (for two-dimensional 'pseudo' PFs) per patient and per optimisation technique ranged from 4 to 17. For a representative example case, the pPFs are shown in figure 3.1.

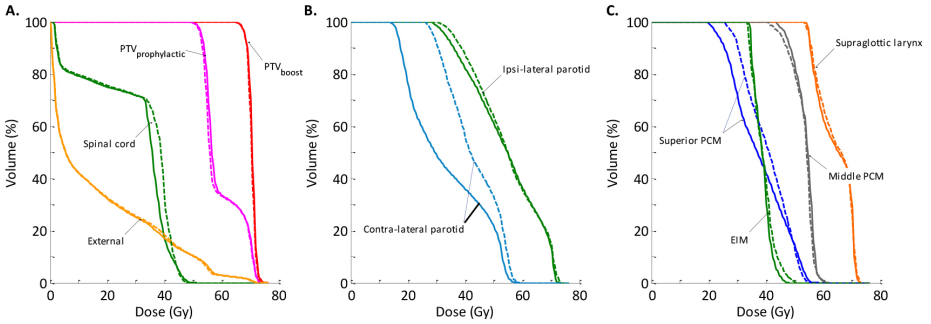
Averaged over all patients, the median of the Pareto optimal sum-NTCP was 5.7% lower (95%CI 1.7–9.9%, p<0.006) for OF_{NTCP}-plans as compared to OF_{gEUD}–plans. The Bland-Altman plots demonstrate that for all patients but one, the sum-NTCP was lower for the OF_{NTCP}-plans (figure 3.2A). No differences were found for the other NTCPs (figure 3.2B-C, table 3.1). Figure 3.2 indicates that an NTCP reduction is feasible, but variable among patients. Furthermore, no relation was found between an NTCP reduction and the location of the primary tumour. It is also shown that for most cases the NTCP for dysphagia could only be improved by deteriorating NTCP for xerostomia and vice versa (figure 3.2D). Other plan evaluation parameters, such as the $CI_{V95\%}$, HI and MUs, and the mean dose to a 1 cm ring around the $PTV_{prophylactic}$, the mean integral dose, and the maximum dose to the spinal cord were not significantly different between both optimisation techniques (table 3.1). Additionally, the results of the manually generated clinical IMRT plans are shown in table 3.1, indicating similar results as the automatically generated plans.

Figure 3.2. *(A-C) Bland-Altman plots for NTCP-values, derived from 'pseudo' Pareto optimal plans, optimised with either a gEUD-based objective function or an objective function based on multivariable NTCP-models.*



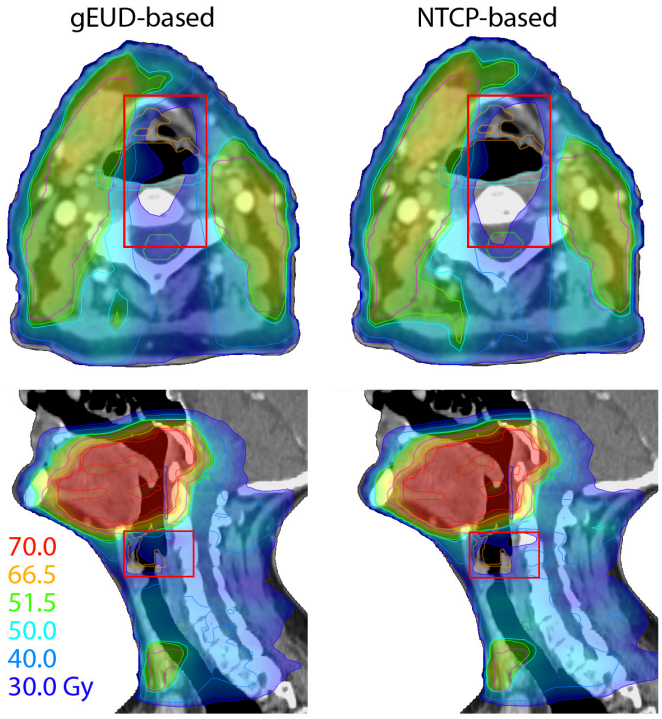
Each data point indicates the difference between the median NTCP estimates per patient and per optimisation technique. Bland-Altman plots of the sum-NTCP (A), NTCP for patient-rated xerostomia (B), and NTCP for physician-rated RTOG grade 2-4 dysphagia (C) are shown. The solid lines and dashed lines represent the mean difference and the 95% (1.96 SD) probability intervals between the data points, respectively. (D) Plot showing the difference in NTCP for dysphagia against the difference in NTCP for xerostomia. The solid line is the regression line.

Figure 3.3. *DVHs of a 'pseudo' Pareto optimal plan (as indicated in Fig. 1) optimised with a gEUD-based objective function (dotted lines) and a plan optimised with an objective function based on multivariable NTCP-models (solid lines).*



(A) DVHs of the targets, dose to the spinal cord and integral dose (External DVH) are similar for the plans. DVHs of salivary dysfunction (B) and swallowing dysfunction (C) related structures. Abbreviations: PTV = planning target volume; EIM = esophagus inlet muscle; PCM = pharlyngeal constrictor muscle.

Figure 3.4. *Transversal and sagittal cross-section for IMRT dose distributions for plans optimised with a gEUD-based objective function (left), and with an objective function based on multivariable normal tissue complication probability models (right) of a patient with a tumour originating in the oropharynx.*



Square boxes indicate differences between the isodose lines at the different organs at risk.

Table 3.1. *Comparison of NTCPs, dosimetric values, and plan evaluation parameters for the (manually generated) clinical IMRT plans and the automatically generated plans.*

	Clinical IMRT-plans
Sum-NTCP (%)	133 (80.8-215)
NTCP patient-rated xerostomia (%)	54.5 (44.0-69.7)
NTCP RTOG grade 2-4 dysphagia (%)	34.9 (7.6-60.2)
NTCP patient-rated swallowing dysfunction (%)	
Solid food	24.1 (8.6-46.1)
Soft food	8.4 (2.1-20.6)
Liquids	6.8 (1.1-12.5)
Choking when swallowing	4.7 (0.9-8.4)
Mean dose OARs (Gy)	
Parotid gland (contra-lateral)	33.3 (22.4-48.4)
Parotid gland (ipsi-lateral)	41.2 (21.8-55.7)
Superior PCM	57.6 (23.0-69.5)
Middle PCM	58.0 (47.6-70.1)
Supraglottic larynx	55.7 (35.7-70.8)
Esophagus inlet muscle	35.8 (23.3-48.2)
Evaluation parameters	
Mean integral dose (external) (Gy)	14.8 (10.2-19.2)
Max dose spinal cord (Gy)	49.7 (47.2-52.3)
Mean dose of 1 cm ring around $PTV_{prophylactic}$ (Gy)	45.0 (43.5-46.5)
$CI_{V95\%}$ ($PTV_{prophylactic}$)	1.5 (1.2-1.7)
$CI_{V95\%}$ (PTV_{boost})	1.3 (1.1-1.4)
HI ($PTV_{prophylactic}$)	0.4 (0.3-0.4)
HI (PTV_{boost})	0.1 (0.1-0.1)
Monitor Units (#)	856 (604-1037)

Abbreviations: NTCP = normal tissue complication probability; OF_{gEUD} = generalised equivalent uniform dose based objective function; OF_{NTCP} = objective function based on multivariable NTCP-models; CI = confidence interval; RTOG = Radiation Therapy Oncology Group; PCM = pharyngeal constrictor muscle; PTV = planning target volume for the boost and prophylactic region; $CI_{V95\%}$ = conformity index of 95% isodose; HI = homogeneity index. The results of the clinical IMRT-plans, the OF_{gEUD} and OF_{NTCP}-plans are expressed as mean (range).

OF_{gEUD}-plans	OF_{NTCP}-plans	Difference (95%CI) [OF_{gEUD} - OF_{NTCP}]	*p*-value
132 (80.3-213)	126 (75.7-204)	5.7 (1.7-9.9)	<0.006
53.8 (48.6-67.5)	52.5 (43.1-64.3)	1.3 (-1.7-4.2)	0.23
34.5 (8.3-61.1)	33.0 (8.8-58.9)	1.5 (-0.4-1.1)	0.08
23.7 (8.8-46.4)	22.4 (9.2-44.1)	1.3 (-0.2-2.8)	0.08
8.3 (2.2-20.8)	8.3 (2.1-20.1)	0.0 (-0.2-0.4)	1.0
6.7 (0.9-12.6)	6.1 (1.0-11.1)	0.6 (0.1-1.1)	0.04
4.6 (0.8-8.4)	4.2 (0.8-7.5)	0.4 (0.1-0.7)	0.04
34.0 (29.5-46.3)	32.9 (24.8-43.3)	1.1 (-1.5-3.7)	0.2
42.2 (25.8-55.8)	41.1 (26.1-55.6)	1.1 (-0.8-3.6)	0.6
57.6 (25.9-69.8)	57.4 (27.4-69.1)	0.2 (-0.6-1.1)	0.3
57.2 (46.7-70.2)	57.4 (46.5-69.6)	-0.2 (-1.0-0.5)	0.9
55.4 (33.2-71.0)	54.3 (34.6-69.1)	1.1 (-0.1-2.4)	0.1
39.4 (34.8-49.7)	38.7 (34.1-47.8)	0.7 (-3.3-3.6)	0.8
14.5 (9.9-18.5)	14.5 (9.8-18.7)	0.0 (-0.5-0.4)	0.7
48.5 (45.7-53.9)	48.4 (46.6-51.9)	0.1 (-1.2-1.6)	0.9
43.2 (41.1-45.2)	42.8 (41.1-44.2)	0.4 (0.0-0.7)	0.1
1.4 (1.2-1.5)	1.4 (1.2-1.5)	0.0 (0.0-0.0)	0.3
1.3 (1.2-1.4)	1.3 (1.2-1.4)	0.0 (0.0-0.1)	0.2
0.4 (0.3-0.4)	0.4 (0.3-0.4)	0.0 (0.0-0.0)	0.6
0.1 (0.1-0.1)	0.1 (0.1-0.1)	0.0 (0.0-0.0)	0.9
827 (720-905)	823 (650-948)	4 (-45-51)	0.8

3.4 Discussion

To our knowledge, this is the first study in which multivariable NTCP-models were incorporated in the objective function for planning optimisation. We have shown that optimisation by convexified NTCP-models is feasible and results in clinically acceptable plans for HNC patients, compared to gEUD-based optimisations. To objectively compare both optimisation techniques, differences were assessed by means of 'pseudo' Pareto front comparisons.

It has been shown that a reduction of the NTCP-values by using OF_{NTCP} compared to using OF_{gEUD} is feasible but relatively small. Apparently, in the studied cases, the gEUD-based optimisations resulted in reasonably well-balanced plans with NTCPs close to those obtained by OF_{NTCP} based optimisations. Importantly, the same NTCP-models were used to both optimise and assess the treatment plans for the OF_{NTCP}-based optimisations, potentially biasing the results in favour of OF_{NTCP}. However, we also showed that the application of OF_{NTCP} resulted in clinically realistic treatment plans with uncompromised target coverage, $CI_{V95\%}$, HI and MU values compared to the gEUD-based plans.

Another reason for the lower (median) NTCP estimates produced by the OF_{NTCP} is due to the implementation of the OF_{gEUD} in Pinnacle[3]. The gEUD-based part of OF_{gEUD} strives for minimization of the relative difference between the $gEUD_j$ (the gEUD for the j-th OAR) and the maximum threshold $gEUD_{0j}$, whereas the NTCP-based part of OF_{NTCP} strives for the minimization of NTCPs. Hence, for OF_{gEUD} there is no incentive to improve $gEUD_j$ below $gEUD_{0j}$. In contrast, for OF_{NTCP}, striving for lower NTCPs is continuously rewarded by a lower value of OF_{NTCP}. This demonstrates the deficiencies of wrapping the gEUD model in a quadratic OF. Therefore, a more appropriate comparison of OF_{gEUD} and OF_{NTCP} would include the implementation of the gEUD-based part in OF_{gEUD} as a 'mean dose' objective, striving for a minimum of the mean dose (without requiring a threshold value).

It was argued by Wu *et al.* that in the gEUD-based optimisation approach fewer planning parameters are required compared to dose-volume-based and other biologically based (e.g. NTCP-based) optimisations [15]. In our study, however, for each gEUD-objective in OF_{gEUD}, a gEUD threshold and weight factor was required. In contrast, for OF_{NTCP}, only a weight factor for each NTCP-model sufficed, as the NTCP-models contain explicit information over the whole dose range, and the additional clinical factors (e.g. age and tumour site) were known prior to plan optimisation. Furthermore, multivariable NTCP-models combine multiple factors into a single objective. A tolerance NTCP threshold as

an inequality constraint in the optimisation problem was not required, as mentioned by Hoffmann *et al.* [30]. The search for an optimal weighting of the planning objectives (either manually or by multi-criteria optimisation techniques) was therefore substantially simplified. In this perspective, we believe that the application of multivariable NTCP-models in the optimisation process provides additional opportunities for automated treatment planning strategies.

A first attempt to optimise treatment plans using biological models was made by Brahme *et al.* introducing P+ as OF, where P+ indicates the complication-free tumor control [31]. Our approach differs from this initial concept by use of a hybrid OF with multiple criteria and the use of multivariable NTCP-models. The emergence of the gEUD-based objectives for planning optimisation was a first practical step in the transition of dose-volume-based to model-based optimisations. Several research groups demonstrated that gEUD criteria for normal tissues improved critical structure sparing in HNC cases, compared to physical dose-volume-based objectives [12,14,15]. However, gEUD objectives remain in the dose domain and do not necessarily correlate linearly with clinical outcome, which is in contrast to direct application of NTCP-models in the optimisation process. The NTCP-models link the treatment plan quantitatively to clinical outcome and shift current treatment approaches towards true individualisation of the treatment. The latter is further supported by the individual prognostic clinical factors included in the models. These variables cause optimal personalisation of the treatment, without increasing the number of subjective weights in the OF. In contrast, the gEUD uses dose information of a single OAR only, and may therefore lead to suboptimal plans.

Our study has some limitations. In general, predictions by empirical models are only accurate if the values of the variables are within the range of values of the data that were used to fit or validate the models. All patient data included in this study were previously used to develop the NTCP-models as presented in this study. Therefore, the NTCP-models were assumed to be valid in this study. In general, validation of the models for the purpose of optimisation is recommended.

For plan evaluation and selection we used the sum of equally weighted NTCP-values. However, it is assumed that a more optimal combination of multiple NTCPs exists which reflects the impact of different complications on a more general OF, e.g., quality of life for the individual patient. This study focused on planning optimisation by means of multivariable NTCP-models as a proof-of-concept. More elaborate translations of individual NTCPs into a meaningful measure will be the subject of ongoing research.

The used NTCP-models comprise a limited number of dose-volume variables. Therefore, the dose in organs not included in one of the NTCP-models may be less restricted than clinically desirable. In our study, this was the case for the ipsilateral parotid, as only the contralateral parotid was incorporated in the NTCP-model for xerostomia. According to the QUANTEC recommendations, however, sparing of one parotid gland may be tolerated when the mean dose to this parotid is <20Gy. Otherwise, the mean doses to both parotids should be <25Gy [32]. Recently, it has been shown that the number of selected variables in a model may increase, while overfitting is still prevented, using more advanced statistical learning methods, such as the least absolute shrinkage and selection operator [33]. It is expected that the inclusion of these NTCP-models in the OF will further improve the global optimisation of the dose distribution.

In conclusion, radiotherapy plan optimisation with direct use of multivariable NTCP-models in the optimisation process is feasible and leads to clinically realistic treatment plans for HNC patients. Furthermore, the use of multivariable NTCP-models facilitates personalised and automated optimisation, reducing the effort to find optimal planning objective settings. These are important steps to reduce radiation-induced side effects and to improve the efficiency of the treatment planning process.

3.5 References

[1] Langendijk JA, Doornaert P, Verdonck-de Leeuw IM, Leemans CR, Aaronson NK, Slotman BJ. Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. J Clin Oncol 2008;26:3770-3776.

[2] Jensen AB, Hansen O, Jorgensen K, Bastholt L. Influence of late side-effects upon daily life after radiotherapy for laryngeal and pharyngeal cancer. Acta Oncol 1994;33:487-491.

[3] Dirix P, Nuyts S. Evidence-based organ-sparing radiotherapy in head and neck cancer. Lancet Oncol 2010;11:85-91.

[4] Chao KS, Majhail N, Huang CJ *et al*. Intensity-modulated radiation therapy reduces late salivary toxicity without compromising tumor control in patients with oropharyngeal carcinoma: a comparison with conventional techniques. Radiother Oncol 2001;61:275-280.

[5] Staffurth J, Radiotherapy Development Board. A review of the clinical evidence for intensity-modulated radiotherapy. Clin Oncol (R Coll Radiol) 2010;22:643-657.

[6] Nutting CM, Morden JP, Harrington KJ *et al*. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. Lancet Oncol 2011;12:127-136.

[7] Ling CC, Li XA. Over the next decade the success of radiation treatment planning will be judged by the immediate biological response of tumor cells rather than by surrogate measures such as dose maximization and uniformity. Med Phys 2005;32:2189-2192.

[8] Allen Li X, Alber M, Deasy JO *et al*. The use and QA of biologically related models for treatment planning: short report of the TG-166 of the therapy physics committee of the AAPM. Med Phys 2012;39:1386-1409.

[9] Kutcher GJ, Burman C. Calculation of complication probability factors for non-uniform normal tissue irradiation: the effective volume method. Int J Radiat Oncol Biol Phys 1989;16:1623-1630.

[10] Niemierko A. A generalized concept of equivalent uniform dose (EUD). Med Phys 1999;26:1100.

[11] Qi XS, Semenenko VA, Li XA. Improved critical structure sparing with biologically based IMRT optimization. Med Phys 2009;36:1790-1799.

[12] Anderson N, Lawford C, Khoo V, Rolfo M, Joon DL, Wada M. Improved normal tissue sparing in head and neck radiotherapy using biological cost function based-IMRT. Technol Cancer Res Treat 2011;10:575-583.

[13] Dirscherl T, Alvarez-Moret J, Bogner L. Advantage of biological over physical optimization in prostate cancer? Z Med Phys 2011;21:228-235.

[14] Hartmann M, Bogner L. Investigation of intensity-modulated radiotherapy optimization with gEUD-based objectives by means of simulated annealing. Med Phys 2008;35:2041-2049.

[15] Wu Q, Djajaputra D, Wu Y, Zhou J, Liu HH, Mohan R. Intensity-modulated radiotherapy optimization with gEUD-guided dose-volume objectives. Phys Med Biol 2003;48:279-291.

[16] Thieke C, Bortfeld T, Niemierko A, Nill S. From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning. Med Phys 2003;30:2332-2339.

[17] Jones L, Hoban P. A comparison of physically and radiobiologically based optimization for IMRT. Med Phys 2002;29:1447-1455.

[18] Semenenko VA, Reitz B, Day E, Qi XS, Miften M, Li XA. Evaluation of a commercial biologically based IMRT treatment planning system. Med Phys 2008;35:5851-5860.

[19] Beetz I, Schilstra C, Burlage FR *et al*. Development of NTCP models for head and neck cancer patients treated with three-dimensional conformal radiotherapy for xerostomia and sticky saliva: the role of dosimetric and clinical factors. Radiother Oncol 2012;105:86-93.

[20] Beetz I, Schilstra C, van der Schaaf A *et al*. NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: the role of dosimetric and clinical factors. Radiother Oncol 2012;105:101-106.

[21] Christianen ME, Schilstra C, Beetz I *et al*. Predictive modelling for swallowing dysfunction after primary (chemo) radiation: results of a prospective observational study. Radiother Oncol 2012;105:107-114.

[22] 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;107:267-273.

[23] van der Laan HP, Christianen ME, Bijl HP, Schilstra C, Langendijk JA. The potential benefit of swallowing sparing intensity modulated radiotherapy to reduce swallowing dysfunction: an in silico planning comparative study. Radiother Oncol 2012;103:76-81.

[24] Christianen ME, Langendijk JA, Westerlaan HE, van de Water TA, Bijl HP. Delineation of organs at risk involved in swallowing for radiotherapy treatment planning. Radiother Oncol 2011;101:394-402.

[25] van de Water TA, Bijl HP, Westerlaan HE, Langendijk JA. Delineation guidelines for organs at risk involved in radiation-induced salivary dysfunction and xerostomia. Radiother Oncol 2009;93:545-552.

[26] Craft DL, Halabi TF, Shih HA, Bortfeld TR. Approximating convex pareto surfaces in multiobjective radiotherapy planning. Med Phys 2006;33:3399-3407.

[27] Janssen T, van Kesteren Z, Franssen G, Damen E, van Vliet C. Pareto fronts in clinical practice for pinnacle. Int J Radiat Oncol Biol Phys 2013;85:873-880.

[28] Hårdemark B, Liander A, Rehbinder H, Löf J. Direct machine parameter optimization with RayMachine in Pinnacle3. RaySearch Laboratories AB 2003.

[29] ICRU. Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT). ICRU Report 83. J ICRU 2010;10:1-106.

[30] Hoffmann AL, den Hertog D, Siem AY, Kaanders JH, Huizenga H. Convex reformulation of biologically-based multi-criteria intensity-modulated radiation therapy optimization including fractionation effects. Phys Med Biol 2008;53:6345-6362.

[31] Brahme A. Optimized radiation therapy based on radiobiological objectives. Semin Radiat Oncol 1999;9:35-47.

[32] Deasy JO, Moiseenko V, Marks L, Chao KS, Nam J, Eisbruch A. Radiotherapy dose-volume effects on salivary gland function. Int J Radiat Oncol Biol Phys 2010;76:S58-63.

[33] Xu CJ, van der Schaaf A, Schilstra C, Langendijk JA, van't Veld AA. Impact of statistical learning methods on the predictive power of multivariate normal tissue complication probability models. Int J Radiat Oncol Biol Phys 2012;82:e677-84.

3.S Supplementary material

3.S.1 NTCP models

Table 3.S1. *Prognostic variables and regression coefficients (β) of the multivariable NTCP-models.*

NTCP-model	Variable	β
Patient-rated moderate to severe xerostomia	Mean dose contralateral parotid gland (Gy)	0.047
	Baseline xerostomia score (none vs. a bit)	0.720
	Constant	-1.443
Physician-rated RTOG grade 2-4 late dysphagia	Mean dose superior PCM (Gy)	0.057
	Mean dose supraglottic larynx (Gy)	0.037
	Constant	-6.09
Problems with swallowing solid food (moderate to severe)	Mean dose superior PCM (Gy)	0.049
	Mean dose supraglottic larynx (Gy)	0.048
	Age (18-65 vs. >65 years)	0.795
	Constant	-6.890
Problems with swallowing soft food (moderate to severe)	Mean dose middle PCM (Gy)	0.061
	Age (18-65 vs. >65 years)	1.203
	Tumour site (other sites vs. oro-/nasopharynx)	1.122
	Radiation technique (3D-CRT vs. IMRT)	-0.912
	Constant	-5.830
Problems with swallowing liquids (moderate to severe)	Mean dose supraglottic larynx (Gy)	0.074
	Radiation technique (3D-CRT vs. IMRT)	-1.209
	Constant	-5.980
Choking when swallowing (moderate to severe)	V60Gy esophagus inlet muscle (%)	0.020
	Mean dose supraglottic larynx (Gy)	0.066
	Constant	-7.07

Abbreviations: NTCP = normal tissue complication probability; RTOG = Radiation Therapy Oncology Group; PCM = pharyngeal constrictor muscle
E.g. the NTCP value for grade 2-4 late dysphagia can be estimated by substitution of S = -6.09 + (mean dose PCM superior x 0.057) + (mean dose supraglottic larynx x 0.037) in Eq. 4 in Supplementary material 3.S2.

3.S.2. Objective functions

In this study, IMRT plans were optimised with two different hybrid objective functions (OFs). Both OFs included physical dose objectives for the PTVs, which were defined as:

$$OF_{PTV} = \sum_{j=1}^{T} \lambda_j \sum_{i=1}^{N} \theta(d_{ij}, d_{0j}) \left(\frac{d_{ij} - d_{0j}}{d_{0j}}\right)^2 v_{ij}, \text{ where} \quad (1)$$

$$\theta(d_{ij}, d_{0j}) = \begin{cases} H(d_{0j} - d_{ij}), & \text{for minimum dose objective} \\ H(d_{ij} - d_{0j}), & \text{for maximum dose objective} \end{cases}$$

and λ_j is a weight factor, d_{ij} the dose in voxel *i*, and v_{ij} the voxel volume relative to the volume of the *j*-th target. H(·) is the Heaviside step function. The criterion is penalized below

or above the dose threshold d_{oj} for minimum and maximum dose objectives, respectively. For the organs at risk (OARs), the first OF (OF_{gEUD}) utilised generalized equivalent uniform dose (gEUD) objectives. In Pinnacle[3] the gEUD is calculated as follows:

$$gEUD = \left(\sum_{i=1}^{N} v_i d_i^a\right)^{1/a} \tag{2}$$

with volume parameter a and other parameters similar to eq. 1.[1]. Now the OF_{gEUD} can be formulated as:

$$OF_{gEUD} = OF_{PTV} + \sum_{j=1}^{M} \lambda'_j \, \theta\left(gEUD_j, gEUD_{0j}\right)\left(\frac{gEUD_j - gEUD_{0j}}{gEUD_{0j}}\right)^2, \text{ where}$$
$$\left(gEUD_j, gEUD_{0j}\right) = \begin{cases} H\left(gEUD_{0j} - gEUD_j\right), & \text{for minimum dose objective} \\ H\left(gEUD_j - gEUD_{0j}\right), & \text{for maximum dose objective} \end{cases} \tag{3}$$

with M objectives for the organs at risk (OARs) with corresponding weight factors λ'_j. The $gEUD$ for the j-th OAR is denoted by $gEUD$ and the corresponding threshold is $gEUD_{oj}$.

For the second OF (OF_{NTCP}) multivariable logistic normal tissue complication probability (NTCP)models were assigned to the OARs. These NTCP-models comprised of n prognostic variables (x) and regression coefficients (β) and were described by the logistic regression formula:

$$\text{NTCP} = \frac{1}{1 + e^{-S}}, \text{ where}$$
$$S = \beta_0 + \sum_{i=1}^{n} \beta_i \cdot x_i \tag{4}$$

During optimisation the NTCP-estimates were only affected by dosimetric changes of the respective OARs, and the logistic NTCP-model is only a function of: 1) the mean dose to the respective OARs (i.e. contralateral parotid gland, superior and middle pharyngeal constrictor muscle, and supraglottic larynx), and the V60Gy of the esophagus inlet muscle. When the OAR dose value exceeds the point of inflection of an NTCP curve, the NTCP is not a convex function of the relevant dose parameter. In this situation, the optimiser may fail to converge to a global minimum [2]. Romeijn *et al*. demonstrated conditions under which non-convex (NTCP-based) OFs used in a multi-criteria (fluence-map) optimisation

problem can be transformed into convex functions while preserving the set of Pareto efficient solutions [3]. Therefore, a strictly increasing transformation (*h*) of equation (4) into a convex function was used:

$$h(\text{NTCP}) = -\ln(1 - \text{NTCP}) \tag{5}$$

A proof for the convexity of similar biological OF is given by Hoffman *et al.*[4]. Note that direct aperture optimisation is a non-convex problem, and the optimisation only benefits from the convex property during fluence-map optimisation. However, in our experience, the optimiser sometimes converged to a better solution using the convexified function. Furthermore, dose-volume parameters (such as V60Gy) are inherently non-convex. These parameters are step functions with respect to voxel dose, with zero derivative nearly everywhere, except at the step boundary, where the derivative is infinite. To avoid infinite values in calculations during steepest descent optimisation the gradient of dose-volume parameters was always set to zero. Hence, the hybrid OF_{NTCP} consisted of OF_{PTV} and *m* NTCP-based objectives, and was formulated as:

$$OF_{NTCP} = OF_{PTV} + \sum_{j=1}^{m} \lambda''_j h(\text{NTCP}_j) \tag{6}$$

where λ''_j is the weight of NTCP_j corresponding to the *j*-th complication endpoint.

3.S.3 References

[1] RaySearch Laboratories AB SS. Biological optimization using the equivalent uniform dose (EUD) in Pinnacle3. RaySearch White Paper. 2003.

[2] Alber M, Nusslin F. An objective function for radiation treatment optimization based on local biological measures. Phys Med Biol 1999;44:479-493.

[3] Romeijn HE, Dempsey JF, Li JG. A unifying framework for multi-criteria fluence map optimization models. Phys Med Biol 2004;49:1991-2013.

[4] Hoffmann AL, den Hertog D, Siem AY, Kaanders JH, Huizenga H. Convex reformulation of biologically-based multi-criteria intensity-modulated radiation therapy optimization including fractionation effects. Phys Med Biol 2008;53:6345-6362.