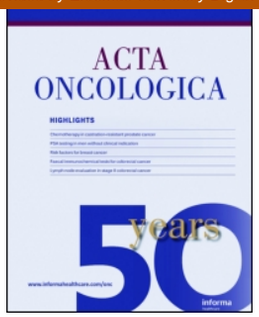

Acta Oncologica

ISSN: 0284-186X (Print) 1651-226X (Online) Journal homepage: http://www.tandfonline.com/loi/ionc20

# Impact of model and dose uncertainty on model-based selection of oropharyngeal cancer patients for proton therapy

Rik G. Bijman , Sebastiaan Breedveld , Tine Arts, Eleftheria Astreinidou, Martin A. de Jong, Patrick V. Granton , Steven F. Petit & Mischa S. Hoogeman

**To cite this article:** Rik G. Bijman , Sebastiaan Breedveld , Tine Arts, Eleftheria Astreinidou, Martin A. de Jong, Patrick V. Granton , Steven F. Petit & Mischa S. Hoogeman (2017): Impact of model and dose uncertainty on model-based selection of oropharyngeal cancer patients for proton therapy, Acta Oncologica, DOI: 10.1080/0284186X.2017.1355113

**To link to this article:** http://dx.doi.org/10.1080/0284186X.2017.1355113

ACTA ONCOLOGICA, 2017
https://doi.org/10.1080/0284186X.2017.1355113

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE



# Impact of model and dose uncertainty on model-based selection of oropharyngeal cancer patients for proton therapy

Rik G. Bijman[a], Sebastiaan Breedveld[a], Tine Arts[a], Eleftheria Astreinidou[b], Martin A. de Jong[b], Patrick V. Granton[a], Steven F. Petit[a] and Mischa S. Hoogeman[a]

[a]Department of Radiation Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands; [b]Department of Radiation Oncology, LUMC, Leiden, The Netherlands

**ABSTRACT**

**Background:** Proton therapy is becoming increasingly available, so it is important to apply objective and individualized patient selection to identify those who are expected to benefit most from proton therapy compared to conventional intensity modulated radiation therapy (IMRT). Comparative treatment planning using normal tissue complication probability (NTCP) evaluation has recently been proposed. This work investigates the impact of NTCP model and dose uncertainties on model-based patient selection.

**Material and Methods:** We used IMRT and intensity modulated proton therapy (IMPT) treatment plans of 78 oropharyngeal cancer patients, which were generated based on automated treatment planning and evaluated based on three published NTCP models. A reduction in NTCP of more than a certain threshold (e.g. 10% lower NTCP) leads to patient selection for IMPT, referred to as 'nominal' selection. To simulate the effect of uncertainties in NTCP-model coefficients (based on reported confidence intervals) and planned doses on the accuracy of model-based patient selection, the Monte Carlo method was used to sample NTCP-model coefficients and doses from a probability distribution centered at their nominal values. Patient selection accuracy within a certain sample was defined as the fraction of patients which had similar selection in both the 'nominal' and 'sampled' scenario.

**Results:** For all three NTCP models, the median patient selection accuracy was found to be above 70% when only NTCP-model uncertainty was considered. Selection accuracy decreased with increasing uncertainty resulting from differences between planned and delivered dose. In case of excessive dose uncertainty, selection accuracy decreased to 60%.

**Conclusion:** Model and dose uncertainty highly influence the accuracy of model-based patient selection for proton therapy. A reduction of NTCP-model uncertainty is necessary to reach more accurate model-based patient selection.

**ARTICLE HISTORY**
Received 8 May 2017
Accepted 5 July 2017



## Introduction

Normal tissue complication probability (NTCP) models have been proposed by Langendijk et al. as a patient-selection tool for new treatment techniques. Their model-based approach uses NTCP-based treatment plan evaluation to analyze NTCP reduction (ΔNTCP) between two alternative techniques. Patients for which ΔNTCP exceeds a certain threshold will be selected for the new treatment technique [1].

Model-based patient selection is particularly useful when outcome of clinical trials is lacking and treatment availability is limited. Intensity modulated proton therapy (IMPT) in head-and-neck cancer is an example of a promising, complication reducing, technique that is more costly and limited in capacity compared to intensity modulated radiation therapy (IMRT) [2–5]. Langendijk et al. and Hoogeman et al. investigated the use of a model-based approach for selection of only those patients that benefit most from IMPT compared to IMRT [1,6].

This paper investigates the currently unknown impact of uncertainty in the used NTCP models and planned dose on selection accuracy of oropharyngeal cancer patients and thereby the clinical usefulness of the model-based approach. NTCP-model uncertainties, represented by confidence intervals (CIs), are included in the publication of the models. Dose uncertainties are generally originating from the differences between planned and delivered dose e.g. induced by anatomy changes during the treatment course [7]. Although these differences are included in the model building, systematic differences between photon and proton treatments may impact the accuracy of patient selection. A Monte Carlo based approach is used to consider the impact of both sources of uncertainty on the accuracy of patient selection.

CONTACT Rik Bijman r.bijman@erasmusmc.nl Department of Radiation Oncology, Erasmus MC Cancer Institute, PO Box 5201, 3008 AE Rotterdam, The Netherlands

Supplemental data for this article can be accessed here.

**Table 1.** Baseline patient characteristics.

| Characteristics | Number | % |
|---|---|---|
| Sex | | |
| Male | 58 | 74 |
| Female | 20 | 26 |
| Age | | |
| <65 | 47 | 60 |
| >65 | 31 | 40 |
| T-classification | | |
| T1 | 4 | 5 |
| T2 | 42 | 54 |
| T3 | 12 | 15 |
| T4 | 20 | 26 |
| Bilateral neck irradiation | | |
| Yes | 71 | 91 |
| No | 7 | 9 |
| Weight loss (before treatment) | | |
| None | 59 | 75 |
| Moderate | 17 | 22 |
| Severe | 2 | 3 |
| Accelerated radiotherapy | | |
| Yes | 38 | 49 |
| No | 40 | 51 |
| Radiotherapy plus Cetuximab | | |
| Yes | 14 | 18 |
| No | 64 | 82 |
| Chemoradiation | | |
| Yes | 22 | 28 |
| No | 56 | 72 |

## Material and methods

### *Patients and treatment planning*

The anonymized CT-scans of 78 oropharyngeal patients treated with IMRT at the Leiden University Medical Center (24 patients) and the Erasmus MC Cancer Institute (54 patients) were used in this study, see Table 1 for the baseline characteristics. Automated treatment planning was used to generate comparable IMRT and IMPT plans to avoid user-dependence on treatment plan quality and enable objective and efficient patient selection [8,9]. The clinical target volumes (CTV) were taken as used clinically. The delineations of the organs at risk were adapted if necessary to comply to the NTCP models used in this study. Dose was prescribed according to a 35 fraction simultaneously integrated boost (SIB) scheme, with a prescription of 70 $Gy_{RBE}$ to the primary tumor and pathological lymph nodes (LUMC) or levels with pathological lymph nodes (Erasmus MC) and 54.25 $Gy_{RBE}$ to the elective nodal areas. For IMPT, minimax robust optimization was used to compensate for rigid and non-rigid patient setup errors, consisting of setup robustness (SR) and range robustness (RR) [10]. Nine error scenarios were considered: six scenarios in the positive and negative direction along three axes, two scenarios simulating a positive and negative range error and one nominal scenario (no errors). Setup error scenarios were simulated by laterally shifting the pencil beams, while range error scenarios were generated by altering the proton energy. For IMRT, margins were added to expand the clinical target volume (CTV) to the planning target volume (PTV). IMRT was planned using 23 equi-angular beams of 6 MV and a dynamic multi-leaf collimator, mimicking a volumetric arc therapy (VMAT) capable dose distribution. For IMPT, a three beams setup (60°, 180° and 300°), with energies ranging from 70 to 250 MeV, was used as suggested by Van de Water et al. [11]. A 57 mm water-equivalent range shifter was used in IMPT to reach target volumes located close to the skin. Constant radiobiological effectiveness (RBE) of 1.0 and 1.1 were assumed for IMRT and IMPT, respectively [12].

Plan generation for IMRT and IMPT was performed fully automatically using our in house developed optimizer, Erasmus-iCycle [8,9]. This optimizer uses the CT scans and corresponding delineations together with a user defined wish-list to automatically generate the treatment plan. The wish-list consists of constraints and prioritized objectives. Constraints should always be met, while objectives are optimized sequentially according to their priorities [13,14]. Comparable wish-lists for IMRT and IMPT plans were designed in close collaboration with radiation oncologists.

The main planning goal was that ≥98% of the PTV (IMRT) did receive ≥95% of the prescribed dose ($V_{95\%} \geq 98\%$) in both the primary tumor and nodal regions. Furthermore, $V_{107\%} \approx 2\%$ and $V_{110\%} \approx 0\%$ should be met for the primary tumor. Also a visual check for hotspots and inconsistencies was performed. For IMPT, the same goals should be met for the CTV of the worst case robustness scenario. Minor violations in target coverage can be corrected for using a rescaling of the dose distribution.

### *Plan comparison*

Plan comparison between 3-mm-margin IMRT plans and 3-mm-SR/3%-RR robustness settings for IMPT plans was performed. This choice for high precision IMRT and IMPT is based on a robustness recipes study by Van der Voort et al. [15] and the expectation that IMPT with 3-mm-SR/3%-RR robustness settings will become available with technical improvements and image-guidance [16]. As a comparison we analyzed other combinations of margins (0, 3 and 5 mm) and SR (0, 3 and 5 mm) and RR (0, 3 and 5%) robustness settings. Note that the margin and robustness settings of choice depend on the desirable and available treatment accuracy.

### *NTCP model-based patient selection*

We used published NTCP models, recently proposed for model-based patient selection in the Netherlands, as a plan comparison and selection tool for IMRT and IMPT treatment as proposed by Langendijk et al. and used by Hoogeman et al. [1,6]. The left part of the flow chart in Figure 1 illustrates the model-based patient selection approach extended with a Monte Carlo method to check for patient selection accuracy. The selection referred to as 'nominal selection' is retrieved according to Langendijk et al. [1], where NTCPs based on IMPT plans are subtracted from those of the IMRT plans, resulting in ΔNTCPs. If the ΔNTCP exceeds a predefined ΔNTCP threshold (e.g. 10 and 5% for grade II and grade III complications, respectively are currently prescribed by the Dutch Society of Radiation Oncology), the patient would be selected for IMPT instead of IMRT. We considered three different NTCP models i.e. for grade II problems swallowing solid food, decreased parotid flow and grade III tube feeding dependence [17–19]. Details regarding the three

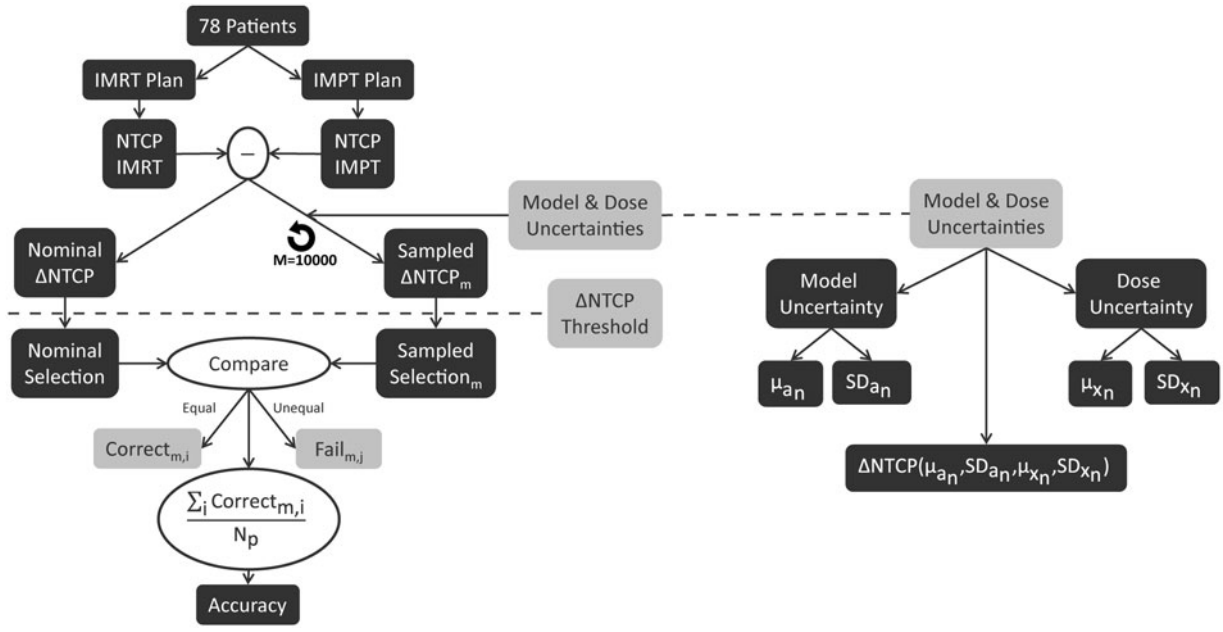


**Figure 1.** Left: Flow chart of the method used. The nominal selection is based on Langendijk et al. [1]. The sampled selection takes into account uncertainty in the NTCP models and the dose. (index i and j represent patients Correct or Failed selected, respectively, while *m* is sample number ranging from 1 to M). $N_p$, refers to the total number of patients. Right: Further elaboration of model and dose uncertainties. Hence, $\mu_{a_n}$ and $\mu_{x_n}$ are the estimates of the nth model coefficients and dose variables respectively, while the uncertainties are introduced via the corresponding standard deviations (SD).

**Table 2.** Details of NTCP models used. The mean values and the confidence intervals of the model coefficients are given.

| Endpoint | Grade | Model type | Model coefficients | Mean | CI |
|---|---|---|---|---|---|
| Problems swallowing food [17] | II | Logistic regression model | Constant | −1.443 | [NA] |
| | | | $D_{\text{superior PCM}}$ | 0.049 | [0.0296–0.0677] |
| | | | $D_{\text{supraglotic larynx}}$ | 0.048 | [0.0100–0.0862] |
| | | | Age | 0.795 | [0.0198–1.5665] |
| < 25% Parotid flow for individual parotid gland after one year [18] | II | Lyman Kutcher Burman (LKB) model | $TD_{50}$ | 39.9 | [37.3–42.8] |
| | | | m | 0.40 | [0.34–0.51] |
| | | | n | 1 | – |
| Tube feeding dependence after six months [19] | III | Logistic regression model | Constant | −11.70 | [-13.47– -8.47] |
| | | | Advanced T-stage | 0.43 | [-0.16–0.73] |
| | | | $\text{Weight loss}_{\text{moderate}}$ | 0.95 | [0.70–1.16] |
| | | | $\text{Weight loss}_{\text{severe}}$ | 1.63 | [1.20–1.99] |
| | | | Accelerated RT | 1.20 | [0.87–1.51] |
| | | | Chemoradiation | 1.91 | [1.39–2.40] |
| | | | RT + Cetuximab | 0.56 | [0.40–0.70] |
| | | | $D_{\text{superior PCM}}$ | 0.071 | [0.044–0.082] |
| | | | $D_{\text{inferior PCM}}$ | 0.034 | [0.006–0.057] |
| | | | $D_{\text{contralateral parotid}}$ | 0.006 | [0–0.0019] |
| | | | $D_{\text{crispharyngeal muscle}}$ | 0.023 | [0.006–0.034] |

$D_x$: mean dose to organ x; RT: radiotherapy; PCM: pharyngeal constrictor muscle; NA: not available.

different NTCP models are provided in Table 2. Patient selections and accuracies were considered for each NTCP model separately.

### *Patient selection accuracy*

To investigate the impact on patient selection subjected to two types of uncertainty, further described below (i.e. NTCP-model uncertainty and dose uncertainty), we drew ten thousand samples from a Gaussian distribution centered at the model coefficient or planned dose values used for the nominal selection. Every sampled ΔNTCP can be seen as an uncertainty-based deviation from the nominal ΔNTCP. Patient selection based on ΔNTCP was done for all samples, after which all sample-based selections were compared to the nominal selection. If a patient is selected for the same modality in the sampled scenario as in the nominal scenario, the selection is considered 'Correct'. For each sampled scenario, the accuracy is the ratio between correctly selected patients and the total number of patients ($N_p = 78$) [20]:

$$\text{Accuracy}_m = \frac{\sum_i \text{Correct}_{m,i}}{N_p}$$

We used the median of all sample accuracies as indicator for the overall accuracy for a specific ΔNTCP threshold and NTCP model. A semi-analytical method, as a statistical more

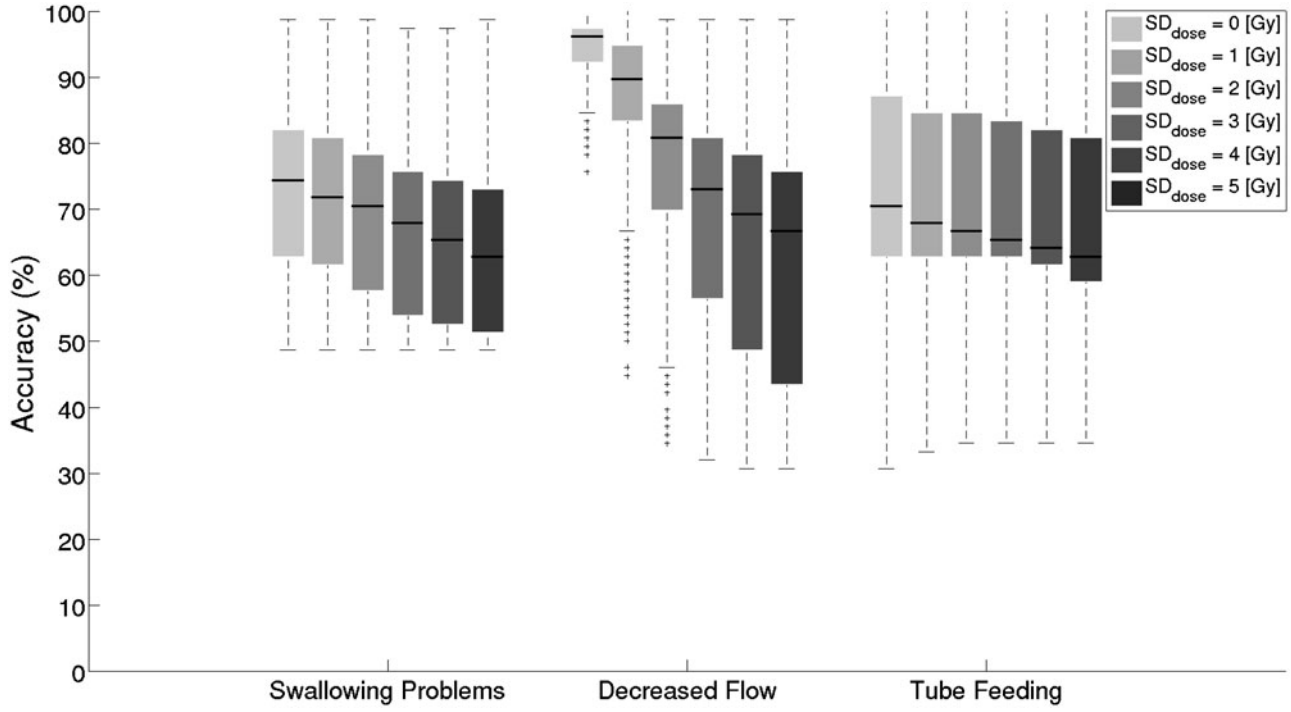


**Figure 2.** Selection accuracy due to combined model and dose delivery uncertainty. The first box, per model, ($SD_{dose} = 0$ Gy) represents the model uncertainty. The subsequent grayscale boxes ($SD_{dose} > 0$ Gy) show the influence of increasing dose delivery uncertainty on the selection accuracy. The median of all samples (2nd quartile) is given by the black solid line, the box around the median defines the 1st and 3rd quartile. The dashed 'whiskers' depict sample accuracies that are not outliers. Outliers (+) are defined as accuracies more than 1.5 times the interquartile distance (size of the box) away from the boxes.



in-to-depth validation of this Monte Carlo method, is elaborated in supplementary material 1.

### NTCP-model uncertainty

Uncertainty in the NTCP models is represented by a probability distribution of the model coefficients ($a_n$) represented by the mean value ($\mu_{a_n}$) and corresponding confidence interval ($CI_{a_n}$). Since our selection accuracy method uses the standard deviation (SD) as a measure of uncertainty, we need to transform the $CI_{a_n}$ into $SD_{a_n}$. According to the central limit theorem, we know that the NTCP-model coefficients originate from a normal probability distribution [21]. This supports the use of the formula for normal distributions to retrieve the standard error (SE) from the CI; $CI = \mu \pm 1.96 \cdot SE$. In case of model coefficient sampling estimates SE is known as SD [21].

### Dose uncertainty

Another source of uncertainty is the difference between planned and actually delivered doses [7]. Those differences are not exactly known but are estimated to be in the order of a few Gy, however, there is evidence for larger differences as well [16,22]. Part of the dose uncertainties are already included in the models' uncertainty as the models have been built from real clinical data. Additionally, systematic dose uncertainties can be of influence on patient selection accuracy. Dose differences due to assumptions of constant RBE, which might be dependent on linear energy transfer (LET) [23] or differences in the impact of anatomy changes between photon and proton treatments are examples of systematic dose uncertainties [7].

We analyzed the effect of increasing systematic dose uncertainty on selection accuracy using dose volume histogram (DVH) samples from a normal distribution centered at the optimized doses ($\mu_{x_n}$) with SDs ranging from no (0 Gy) to excessive (5 Gy) uncertainty in the doses ($SD_{x_n}$) [24].

In the results section we will make a distinction between the impact of uncertainty in the NTCP models and the additional impact of systematic dose uncertainty. This is shown because model uncertainty is a known and published source of uncertainty, while the actual uncertainty in the dose is not.

### Influence of ΔNTCP threshold, model improvement and dose correlation

As a further consideration we investigated the sensitivity of selection accuracy with respect to different ΔNTCP thresholds. We also analyzed by how much selection accuracy could be increased when uncertainty in the NTCP models is reduced. This was done by means of a hypothetical reduction of the model coefficient CIs by factor 2. Furthermore, the influence of correlation between doses in different organs at risk on selection accuracy was taken into consideration. Correlation between dose uncertainties was introduced by changing the dose sampling strategy from independent to dependent.

## Results

Figure 2 shows that in all three models the median accuracy level (solid black line) was above 70% based on model

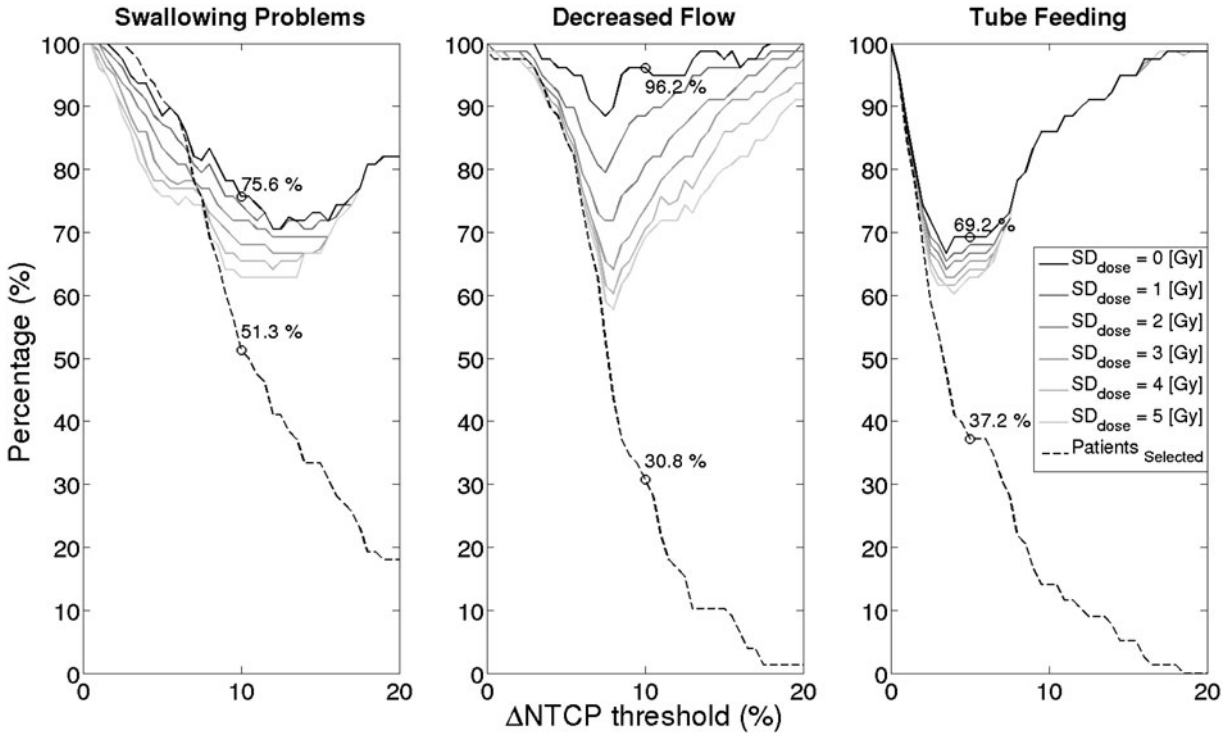


**Figure 3.** Selection accuracy due to model coefficient (SD = 0 Gy)and dose ($SD_{dose} > 0$ Gy) uncertainty (solid) as a function of ΔNTCP threshold. Patient selection rate (dashed) shows impact of ΔNTCP threshold increase on patient selection. The numbers represent the values at the prescribed ΔNTCP.

**Table 3:** Accuracy as function of margin (IMRT, columns) and robustness settings (IMPT, rows) at prescribed ΔNTCP threshold and $SD_{dose} = 3$ Gy.

| | Swallowing problems | | | Decreased flow | | | Tube feeding | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0 mm | 3 mm | 5 mm | 0 mm | 3 mm | 5 mm | 0 mm | 3 mm | 5 mm |
| 0 mm–0% | 68 | 78 | 87 | 81 | 81 | 87 | 74 | 67 | 68 |
| 3 mm–3% | 77 | 68 | 81 | 97 | 74 | 80 | 86 | 65 | 64 |
| 3 mm–5% | 76 | 67 | 80 | 99 | 77 | 76 | 90 | 65 | 64 |
| 5 mm–3% | 86 | 68 | 72 | 100 | 91 | 73 | 95 | 68 | 62 |
| 5 mm–5% | 85 | 69 | 69 | 100 | 94 | 74 | 95 | 68 | 63 |

uncertainty alone. Increasing dose uncertainty resulted in the largest accuracy drop for the decreased parotid flow model. The median accuracy level, however, never decreased below 60% even for excessive dose uncertainties. Compared to dose uncertainty, model uncertainty was most important in the models for problems swallowing solid food and tube feeding dependence. The accuracy spread was stable for the swallowing problems and tube feeding models, while the spread increased with dose uncertainty for the model of decreased parotid flow.

Figure 3 shows that median accuracy levels at the currently prescribed ΔNTCP thresholds for grade II (III) of 10% (5%) were near the lowest point on the curves and additional reduction of selection accuracy with increasing dose uncertainty was most severe.

Accuracy of patient selection based on plan comparison of 3-mm-margin against 3-mm-SR/3%-RR had been considered so far. Hence, Table 3 shows that there was no general trend for accuracy level as function of margin or robustness setting. However, the highest accuracy levels seemed to occur when the differences between robustness and margin increased. The same trend was found at different levels of dose uncertainties.

A hypothetical model improvement regarding uncertainty in NTCP models resulted in a model accuracy increase of up to more than 10%. The accuracy levels due to model uncertainty only increased to more than 80% for all three models. The same increase was found for the combined accuracy at a moderate dose uncertainty level ($SD_{dose} = 3$ Gy). The largest gain in selection accuracy could be achieved for the models of swallowing problems and tube feeding.

## Discussion

In this study we investigated the impact of model and dose variable uncertainties on the accuracy of patient selection for IMPT. Patient selection based on sampled ΔNTCPs, representing uncertainties in model coefficients and doses, were compared to nominal selections. We found median selection accuracies to range from 63% to 96%. For swallowing problems and tube feeding dependence the selection accuracy was most affected by the uncertainty in the model coefficients. Contrary for the model of decreased parotid flow, which was most effected by uncertainty in dose.

The relatively low impact of model uncertainties on the selection accuracy for the decreased parotid flow model could be explained by the fact that only two coefficients were included in the model. Moreover, the two coefficients were fitted to data of 222 patients and in total 384 parotid glands from two independent cohorts [18]. The models for swallowing problems and tube feeding contained 4 and 11 coefficients respectively, which were fitted using data of 354

and 355 patients [17,19] . Generally speaking, if more variables are included in multi-variable modeling, more patients (with events) need to be included for an accurate estimate of the model coefficients.

As a proof of concept we analyzed what impact a reduction in the uncertainty of the model coefficients could have. The results showed that for swallowing problems and tube feeding the selection accuracy could be improved to above 80%. Based on these findings, we recommend the inclusion of more patients in multi-variable NTCP modeling in order to lower the CIs.

The model for decreased parotid flow showed a large dependence on additional uncertainties in dose. This could be explained by the fact that NTCP values were on average located in a steeper part of the NTCP curve. This makes the selection according to this model inherently more sensitive to uncertainties in dose. A second reason for the lower impact of dose uncertainty on patient selection in the models for swallowing problems and tube feeding could be that complication is predicted by multiple dose variables. In this study, dose samples were drawn independently, while they may be correlated. Therefore, we also investigated possible correlation between different dose uncertainties, but this did not lead to large differences in accuracy. Selection accuracy also depends on the ΔNTCP threshold. The minimum thresholds as prescribed by the Dutch Society of Radiation Oncology resulted in suboptimal accuracy levels (see figure 3). For the tube feeding model the accuracy can be increased to above 80% by increasing the ΔNTCP threshold from 5% to 10%, however this will also lead to less patients being selected for proton therapy. Reduction of selection accuracy with increasing dose uncertainty appeared to be most severe in the region of the currently proposed ΔNTCP thresholds, which underpins the importance of keeping the difference between planned and delivered dose as low as possible.

In this study we compared 3-mm margin IMRT with 3-mm-SR/3%-RR robust IMPT. We would like to stress that the IMRT margins and IMPT robustness settings that would result in adequate target coverage depend on the level of dose delivery accuracy that can be achieved with each of the modalities. Therefore, we also determined the selection accuracy for other combinations of margins and robustness settings at different dose uncertainty levels. We found the lowest accuracy for combinations of comparable margins and robustness settings. Larger margins lead to increasing mean doses and thus higher NTCPs for the IMRT plan. As long as the inflection point of the NTCP-curve is not reached, this will lead to larger ΔNTCPs that are further away from the threshold. This increases the correct selection rate and thereby the selection accuracy. Accuracy increase due to larger robustness settings can be explained using a similar reasoning.

The main focus of this study was to investigate the influence of uncertainties on patient selection accuracy. Note that the model-based approach will not select patients for whom a complication will be avoided if they are treated with proton therapy. In this regard, it might be interesting to calculate other metrics, which focus on the impact of uncertainties on the expected ΔNTCP for patients selected for a photon or proton treatment. For example, the ratio of sampled ΔNTCP and nominal ΔNTCP values can be calculated for this purpose.

The dose uncertainty analysis focused on the impact of systematic dose uncertainties (e.g. LET depended RBE) assuming that the random dose uncertainties are already incorporated in the model uncertainty. However, one could think of random dose uncertainties that were not present in the NTCP model building cohort. An example is the difference between the dose distribution used for patient selection and that in the final deliverable treatment plan. These differences can occur if predicted organ-at-risk DVHs, are used for patient selection, which was proposed by Delaney et al. [22]. In our study, we used fully generated treatment plans for patient selection to minimize the dose uncertainty.

A limitation of this study is that we assumed that NTCP models derived from photon treatments are also valid for IMPT. Recently, Blanchard et al. were the first that validated photon derived NTCP models for patients treated with proton therapy [25]. They found that the models remained valid suggesting that this source of uncertainty can be ignored compared to uncertainties in model coefficients and dose. Nevertheless, as soon as model-based patient selection is used in clinical practice, treatment-related toxicities should be captured prospectively in order to validate the NTCP models for proton therapy. Another remark is the fact that information about confidence intervals was missing for the constant in the swallowing problems model. For this reason the actual selection accuracy due to uncertainty in NTCP models might be somewhat lower for this model.

The guidelines of the Dutch Society of Radiation Oncology allow the addition of ΔNTCPs from different models before making the selection. It would be interesting to investigate if certain combinations would lead to improved selection accuracy.

We showed that the accuracy of model-based patient selection for proton therapy is highly affected by uncertainties in NTCP models and uncertainties resulting from differences between planned and delivered dose variables. Since accurate patient selection for IMPT is of utmost importance, we recommend that the uncertainty in NTCP models is reduced by analyzing sufficiently large patient cohorts and by external validation of the models. Furthermore, since the reduction of selection accuracy appeared to be most severe in the region of the currently proposed ΔNTCP thresholds, the difference between planned and delivered doses should be kept as low as possible.

## Disclosure statement

Erasmus MC Cancer Institute has research collaborations with Elekta AB, Stockholm, Sweden and Accuray Inc, Sunnyvale, USA and Varian Medical Systems, Palo Alto, USA.

## ORCID

Rik G. Bijman http://orcid.org/0000-0003-1915-3739

Sebastiaan Breedveld http://orcid.org/0000-0001-8954-4554
Patrick V. Granton http://orcid.org/0000-0002-2635-2894
Mischa S. Hoogeman http://orcid.org/0000-0002-4264-9903

## References

[1] Langendijk J, Lambin P, De Ruysscher D, et al. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. Radiother Oncol. 2013;107:267–273.
[2] Ares C, Hug EB, Lomax AJ, et al. Effectiveness and safety of spot scanning proton radiation therapy for chordomas and chondrosarcomas of the skull base: first long-term report. Int J Radiat Oncol Biol Phys. 2009;75:1111–1118.
[3] Rutz HP, Weber DC, Goitein G, et al. Postoperative spot-scanning proton radiation therapy for chordoma and chondrosarcoma in children and adoloscents: initial experience at Paul Scherrer insitute. Int J Radiat Oncol Biol Phys. 2008;71:220–225.
[4] Schultz-Ertner D, Nikoghosyan A, Thilmann C, et al. Results of carbon ion radiotherapy in 152 patients. Int J Radiat Oncol Biol Phys. 2004;58:631–640.
[5] Jakobi A, Bandurska-Lugua A, Stützer K, et al. Identification of patient benefit from proton therapy for advanced head and neck cancer patients based on individual and subgroup normal tissue complication probability analysis. Int J Radiat Oncol Biol Phys. 2015;92:1165–1174.
[6] Hoogeman M, Breedveld S, De Jong M, et al. Proton therapy patient selection for oropharyngeal cancer patients: the impact of treatment accuracy. Radiother Oncol 2017;123:OC-0154.
[7] Vásquez Osorio EM, Hoogeman MS, Al-Mamgani A, et al. Local anatomic changes in parotid and submandibular glands during radiotherapy for oropharynx cancer and correlation with dose, studied in detail with nonrigid registration. Int J Radiat Oncol Biol Phys. 2008;70:875–882.
[8] Breedveld S, Storchi PR, Voet PW, et al. I-Cycle: integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans. Med Phys. 2012;39:951–963.
[9] Breedveld S, Storchi PR, Heijmen BJ. The equivalence of multi-criteria methods for radiotherapy plan optimization. Phys Med Biol. 2009;54:7199–7209.
[10] Frederiksson A, Forsgren A, Hårdemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. Med Phys. 2011;38:1672–1684.
[11] Van de Water TA, Lomax AJ, Bijl HP, et al. Using a reduced spot size for intensity modulated proton therapy potentially improves salivary gland-sparing in oropharyngeal cancer. Int J Radiat Oncol Biol Phys. 2012;82:313–319.
[12] Paganetti I, Niemierko A, Ancukiewicz M, et al. Relative biological effectiveness (RBE) values for proton beam therapy. Int J Radiat Oncol Biol Phys. 2002;53:407–421.
[13] Voet PW, Dirkx ML, Breedveld S, et al. Toward fully automated multicriterial plan generation: a prospective clinical study. Int J Radiat Oncol Biol Phys. 2013;85:866–872.
[14] Van de Water S, van Dam I, Schaart DR, et al. The price of robustness; impact of worst-case optimization on organ-at-risk in intensity-modulated proton therapy for oropharyngeal cancer patients. Radiother Oncol. 2016;120:56–62.
[15] Van der Voort S, van de Water S, Perkó Z, et al. Robustness recipes for minimax robust optimization in intensity modulated proton therapy for oropharyngeal cancer patients. Int J Radiat Oncol Biol Phys. 2016;95:163–170
[16] Kraan A, van de Water S, Teguh D, et al. Dose uncertainties in IMPT for oropharyngeal cancer in the presence of anatomical, range and setup errors. Int J Radiat Oncol Biol Phys. 2013; 87:888–896.
[17] 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.
[18] Dijkema T, Raaijmakers CP, Ten Haken RK, et al. Parotid gland function after radiotherapy: the combined Michigan and Utrecht experience. Int J Radiat Oncol Biol Phys. 2010;78:449–453.
[19] Wopken K, Bijl HP, van der Schaaf A, et al. Development of a multivariable normal tissue complication probability (NTCP) model for tube feeding dependence after curative radiotherapy/chemoradiotherapy in head and neck cancer. Radiother Oncol. 2014;113:95–101.
[20] Fawcett T. An introduction to ROC analysis. Patern Recogn Lett. 2006;27:861–874.
[21] Agresti A, Finlay B. Statistical methods for the social sciences. 4th ed. Upper Saddle River: Pearson International Edition; 2009.
[22] Delaney A, Dahele M, Tol J, et al. Using a knowledge-based planning solution to select patients for proton therapy. Radiother Oncol . Forthcoming. 2017. DOI:10.1016/j.radonc.2017.03.020
[23] Giulia G, Böhlen T, Cabal G, et al. Variable RBE in proton therapy: comparison of different model predictions and their influence on clinical-like scenarios. Radiat Oncol. 2016;11:68–83.
[24] Astreinidou E, Bel A, Raaijmakers C, et al. Adequate margins for random setup uncertainties in head-and-neck IMRT. Int J Radiat Oncol Biol Phys. 2005;61:938–944.
[25] Blanchard P, Wong AJ, Gunn GB, 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;121:381–386.