Prediction and prevention of radiation-induced swallowing dysfunction
Christianen, Miranda Eligia Maria Cornelia

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

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

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

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

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

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

Download date: 21-10-2017
CHAPTER 5
Swallowing sparing intensity modulated radiotherapy (SW-IMRT) in head and neck cancer: validation according to the model-based approach


Accepted for publication Radiotherapy and Oncology (2015)
CHAPTER 5

Abstract

Purpose: The aim of this study was to clinically validate a multivariable normal tissue complication probability (NTCP) model for grade 2-4 swallowing dysfunction at 6 months after radiotherapy or chemoradiation (SWALM6) in head and neck cancer patients treated with swallowing sparing intensity modulated radiotherapy (SW-IMRT) and to test if SW-IMRT resulted in reduction of the prevalence of SWALM6.

Materials and methods: The primary endpoint was SWALM6. For all 186 patients, a standard IMRT (parotid sparing) and a SW-IMRT plan (additional constraints for swallowing organs at risk) was created. The difference in NTCP for SWALM6 (ΔNTCP_{SWALM6} = NTCP_{standard} – NTCP_{SW-IMRT}) was calculated. Patients were treated with SW-IMRT. The external validation of the NTCP model was analyzed by comparing performance measures.

Results: The mean ΔNTCP_{SWALM6} was 4.9% (range 0.01 – 17.3%), with a significant lower mean predicted NTCP_{SW-IMRT} of 22.6% (95% CI 20.2 – 24.9%), compared to NTCP_{standard} of 27.5% (95% CI 24.9 – 29.9%) (P<.001). There was a perfect correspondence of NTCP_{SW-IMRT} with the observed prevalence of SWALM6 (22.6%). The overall model performance, discrimination and ‘goodness of fit’ were good.

Conclusion: We externally validated the multivariable NTCP model for SWALM6 in SW-IMRT treated patients, showing reduced swallowing dysfunction by reducing the dose parameters included in this NTCP model.
Introduction

Swallowing dysfunction is one of the most devastating side effects after definitive radiotherapy (RT) or chemoradiation (CHRT) for head and neck cancer (HNC) and has a major impact on health-related quality of life (HRQoL). Recently, we reported on the results of a large multicenter prospective cohort study in which we developed multivariable Normal Tissue Complication Probability (NTCP) models for swallowing dysfunction. In that study we identified two independent risk factors for grade 2-4 swallowing dysfunction at 6 months after completion of treatment (SWALM6), including the mean dose to the superior pharyngeal constrictor muscle (superior PCM) and the mean dose to the supraglottic larynx. We subsequently showed that swallowing sparing intensity modulated radiotherapy (SW-IMRT) is expected to result in clinically relevant reductions in the risk of swallowing dysfunction in approximately half of the patients. SW-IMRT refers to IMRT with dose constraints for both the parotid glands as well as for the swallowing organs at risk (SWOARs), without compromising the dose to the planning target volumes (PTV) and the parotid glands. However, the clinical validation of SW-IMRT remains to be determined.

Therefore, the purpose of this study was to clinically validate a previously developed multivariable NTCP model for SWALM6 in a cohort of patients treated with SW-IMRT and if this resulted in a reduction of the prevalence of SWALM6 as compared to standard IMRT (parotid gland sparing IMRT).

Methods and Materials

Patients

The study population of this prospective cohort study was composed of 186 consecutive patients treated from September 2010 to September 2014 at the Department of Radiation Oncology of two medical centers in the Netherlands: the VU University Medical Center (VUMC), Amsterdam or the University Medical Center Groningen (UMCG), Groningen. All patients were treated with definitive RT for squamous cell HNC originating from the oral cavity, pharynx or larynx, either alone or in combination with concomitant chemotherapy or cetuximab. All patients were subjected to a standard follow-up program that included prospective evaluation of acute and late toxicity, patient-rated symptoms and HRQoL, prior to, during and at regular intervals after treatment.

Patients who previously underwent surgery, radiotherapy and/or chemotherapy and those with prior malignancies, and/or distant metastases were excluded. Patients with RTOG grade 2-4 swallowing dysfunction at baseline were also excluded in order to ensure that the observed swallowing dysfunction was induced by radiation treatment itself and not by tumor extension. Furthermore patients with residual disease or recurrence within 6 months were excluded.
### Table 1: Patients characteristics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>All patients (N=186)</th>
<th>LOW NTCP reduction group (ΔNTCP: 0-5%) (N=99)</th>
<th>HIGH NTCP reduction group (ΔNTCP &gt;5%) (N=87)</th>
<th>P-value *</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>%</td>
<td>Number</td>
<td>%</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>138</td>
<td>74</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>48</td>
<td>26</td>
<td>22</td>
</tr>
<tr>
<td>Age, years</td>
<td>18-65</td>
<td>109</td>
<td>59</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>&gt; 65</td>
<td>77</td>
<td>41</td>
<td>44</td>
</tr>
<tr>
<td>Tumor classification</td>
<td>T1-T2</td>
<td>103</td>
<td>55</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>T3-T4</td>
<td>83</td>
<td>45</td>
<td>36</td>
</tr>
<tr>
<td>Node classification</td>
<td>N0</td>
<td>96</td>
<td>52</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>N+</td>
<td>90</td>
<td>48</td>
<td>42</td>
</tr>
<tr>
<td>Primary Site</td>
<td>Larynx</td>
<td>85</td>
<td>46</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Oropharynx</td>
<td>64</td>
<td>35</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Oral cavity</td>
<td>8</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Hypopharynx</td>
<td>21</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Nasopharynx</td>
<td>8</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Treatment modalities</td>
<td>Conventional RT</td>
<td>51</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Accelerated RT</td>
<td>63</td>
<td>34</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Chemoradiation</td>
<td>61</td>
<td>33</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Cetuximab with RT</td>
<td>11</td>
<td>6</td>
<td>9</td>
</tr>
</tbody>
</table>

* P-value LOW NTCP reduction versus HIGH NTCP reduction, based on chi-square. Abbreviations: N=number, NTCP=normal tissue complication probability, RT=radiotherapy

### Endpoints

The primary endpoint was defined as grade 2-4 swallowing dysfunction according to the RTOG/EORTC Late Radiation Morbidity Scoring Criteria, as assessed 6 months after completion of treatment (SWALM6), which means that patients were not able to eat solid food and were only able to eat semi-solid food, swallow liquids or were dependent on tube feeding.

### Treatment

All organs at risk (OARs), including the salivary glands, and the SWOARs were delineated as previously described\(^{10,11}\). Regions of interest, IMRT planning and optimization for SW-IMRT were described previously in detail\(^8\).

### Study design and statistical analysis

The validation of SW-IMRT was performed according to the model-based approach\(^12\). Model-based validation means that multivariable NTCP models developed in a population with a certain radiation technique (in this case standard IMRT and three-dimensional conformal RT (3D-CRT)) are also valid in an independent subsequent population treated with another radiation technique (in this case SW-IMRT) (external validation). This implies that dose reductions (ΔDose) obtained with the new technology results in NTCP reductions...
(ΔNTCP) as predicted by the multivariable NTCP model and thus that the new technique indeed contributes to less toxicity. For this purpose, the following steps were made:

Step 1 (NTCP\text{standard}): For all patients, a standard IMRT treatment plan was created, using dose constraints for the parotid glands but without dose constraints for the SWOARs. For this plan, the NTCP value for SWAL\text{M6} (NTCP\text{standard}) was calculated using the equitation of the prediction model for SWAL\text{M6} as described by Christianen et al\textsuperscript{7}:

\[ NTCP = \frac{1}{1 + e^{-S}} \]

in which \( S = -6.09 + (\text{mean dose superior PCM} \times 0.057) + (\text{mean dose supraglottic larynx} \times 0.037) \). This plan was then saved and stored.

Step 2 (NTCP\text{SW-IMRT}): This standard IMRT plan was then further optimized into a SW-IMRT plan with similar planning objectives for the parotid glands and target volumes, but with additional constraints for the SWOARs. During the planning procedures, attempts were made to reduce the dose to the SWOARs as much as possible in an iterative process, until the dose to other organs at risk started to rise (parotid glands, oral cavity or spinal cord) and/or the dose to the PTV’s were compromised. For this plan, the NTCP value for SWAL\text{M6} (NTCP\text{SW-IMRT}) was calculated with the same equation that was used in step 1.

Subsequently, ΔNTCP SWAL\text{M6} was calculated, defined as NTCP\text{standard} – NTCP\text{SW-IMRT}, and thus corresponding with the predicted NTCP reduction that could be obtained with SW-IMRT as compared to standard IMRT in each individual patient. Patients were divided into 2 groups, including a LOW ΔNTCP\text{SWAL\text{M6}} group if ΔNTCP\text{SWAL\text{M6}} was 0-5% and a HIGH ΔNTCP\text{SWAL\text{M6}} group if ΔNTCP\text{SWAL\text{M6}} was more than 5%. This threshold was chosen in advance on an arbitrary basis.

Step 3 (Actual treatment): Patients were then actually treated with SW-IMRT and prospectively followed using exactly the same data registration program as used in the previous population treated with 3D-CRT or standard IMRT in which the multivariable model was developed\textsuperscript{7}.

Step 4 (External model validation): The external validation of the NTCP model was done by a number of performance measures of the NTCP model in the SW-IMRT cohort to that obtained in the original cohort treated with standard IMRT or 3D-CRT that was used to develop the model\textsuperscript{7}. For this purpose, we used Monte-Carlo simulations using repeated random drawings of the outcomes according to the model NTCP to generate the expected distributions of the performance measures, based on the model and the case-mix of the cohort, and calculated single sided p-values (i.e., testing the null-hypothesis that the actual performance was not worse than expected). Overall model performance was described with the explained variance (using the Nagelkerkes R\textsuperscript{2}) and the scaled Brier score. The Brier score is the average squared difference between the predicted probability and the actual outcome. The scaled Brier score is a recalculated Brier score that will give a more robust comparability of the accuracy of the model. A scaled Brier score should be as close to 1 as possible (a perfect model) and is 0 for a non-informative model\textsuperscript{13,14}. For the discrimination ability of the model, we calculated the area under the receiver operating characteristics curve (AUC), and the discrimination slope, defined as the absolute dif-
Chapte 5

dherence between the mean predicted NTCP value for patients with and without the out-

come\textsuperscript{13,14}. Finally, a Hosmer-Lemeshow goodness-of-fit test was performed to evaluate
the calibration of the model. The model’s predictions fit the data at an acceptable level if
the Hosmer-Lemeshow goodness-of-fit test statistic is >0.05\textsuperscript{15}. Since in the previous co-
hort\textsuperscript{7} we only assessed model performance using the AUC, we retrospectively performed
all the above mentioned model performance measures in the previous cohort as well.

Step 5 (Technique validation): To evaluate the ability of SW-IMRT to reduce the prev-
elence of SWALM\textsuperscript{6}, we analyzed if the observed prevalence of SWALM\textsuperscript{6} significantly dif-
fered from the average predicted NTCP\textsubscript{standard}, using a non-parametric two-sided test.

Data were analyzed using SPSS Statistics for Windows, version 19.

Results

Patient characteristics
The study population of this prospective cohort study was composed of 186 patients with
a mean age of 64 years. The majority of patients were male (74%). The pre-treatment
characteristics of the patients are listed in Table 1.

Dose distribution in organs at risk
The mean dose in all SWOARs and the contralateral parotid glands was significantly lower
with the SW-IMRT plans compared to the standard IMRT plans, whereas the mean dose
in the ipsilateral parotid and submandibular glands remained the same. (Table 2A).

\begin{table}[h!]
\centering
\caption{Dose distribution parameters}
\begin{tabular}{lcccc}
\hline
Organ at risk & Dose (Gy) according to actually given IMRT plan & Dose (Gy) according to BACK UP standard IMRT plan & Difference between SW-IMRT and Standard IMRT (Gy) & P-value \textsuperscript{a} \\
\hline
\textbf{A} Superior pharyngeal constrictor muscle & 41.5 & 38.2-44.7 & 44.4 & 41.0-47.7 & 2.9 & < 0.001 \\
Middle pharyngeal constrictor muscle & 46.8 & 43.9-49.7 & 50.9 & 48.2-53.5 & 4.1 & < 0.001 \\
Inferior pharyngeal constrictor muscle & 53.7 & 51.4-55.7 & 57.4 & 55.4-59.2 & 3.7 & < 0.001 \\
Cricopharyngeal muscle & 49.2 & 47.1-51.2 & 52.2 & 50.2-54.0 & 3.0 & < 0.001 \\
Esophageal inlet muscle & 35.4 & 33.0-37.5 & 41.7 & 39.2-43.7 & 6.3 & < 0.001 \\
Supraglottic larynx & 54.3 & 52.0-56.3 & 58.1 & 56.1-59.8 & 3.8 & < 0.001 \\
Glottic larynx & 56.6 & 54.0-58.9 & 58.9 & 56.7-61.0 & 2.3 & < 0.001 \\
Parotid gland ipsilateral & 27.8 & 25.4-30.4 & 27.9 & 25.6-30.4 & 0.1 & 0.120 \\
Parotid gland contralateral & 20.9 & 19.1-22.7 & 21.5 & 19.7-23.3 & 0.6 & < 0.001 \\
Submandibular gland ipsilateral & 49.6 & 46.0-53.1 & 49.8 & 46.2-53.2 & 0.2 & 0.044 \\
Submandibular gland contralateral & 44.0 & 40.5-47.2 & 44.3 & 40.8-47.5 & 0.3 & 0.003 \\
\hline
\end{tabular}
\end{table}
## ΔNTCP

The ΔNTCP\textsubscript{SWALM6}, varied widely between individual patients, with a mean ΔNTCP\textsubscript{SWALM6} of 4.9% (range 0.01 to 17.3%). Out of 186 patients, 87 (47%) were classified as having HIGH ΔNTCP\textsubscript{SWALM6} while 99 patients (53%) were classified as having LOW ΔNTCP\textsubscript{SWALM6} (Figure 1). The LOW and HIGH ΔNTCP\textsubscript{SWALM6} groups differed significantly with regard to a number of pretreatment variables (Table 1). The patients in the HIGH ΔNTCP\textsubscript{SWALM6} group had higher T-stages, more primary tumors originating from nasopharynx and oropharynx, and were more often treated with CHRT or conventional RT (Table 1). Moreover, the doses delivered to some of the SWOARs were significantly higher (Table 2B and Table 2C).

### Table 1

<table>
<thead>
<tr>
<th>Organ at risk</th>
<th>Dose (Gy) according to actually given SW-IMRT and Standard IMRT</th>
<th>P-value *</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average Dmean</td>
<td>95% CI</td>
</tr>
<tr>
<td>Superior pharyngeal constrictor muscle</td>
<td>34.3</td>
<td>29.0-39.4</td>
</tr>
<tr>
<td>Middle pharyngeal constrictor muscle</td>
<td>41.0</td>
<td>36.2-45.8</td>
</tr>
<tr>
<td>Inferior pharyngeal constrictor muscle</td>
<td>54.9</td>
<td>51.5-57.9</td>
</tr>
<tr>
<td>Cricopharyngeal muscle</td>
<td>48.7</td>
<td>45.6-51.8</td>
</tr>
<tr>
<td>Esophageal inlet muscle</td>
<td>30.7</td>
<td>27.1-34.1</td>
</tr>
<tr>
<td>Supraglottic larynx</td>
<td>54.2</td>
<td>50.4-57.1</td>
</tr>
<tr>
<td>Glottic larynx</td>
<td>59.1</td>
<td>55.6-62.4</td>
</tr>
<tr>
<td>Parotid gland ipsilateral</td>
<td>21.9</td>
<td>18.4-25.3</td>
</tr>
<tr>
<td>Parotid gland contralateral</td>
<td>16.5</td>
<td>13.7-19.2</td>
</tr>
<tr>
<td>Submandibular gland ipsilateral</td>
<td>39.9</td>
<td>34.1-45.5</td>
</tr>
<tr>
<td>Submandibular gland contralateral</td>
<td>34.7</td>
<td>29.2-39.6</td>
</tr>
</tbody>
</table>

* P-value actually given dose SW-IMRT versus BACK UP standard IMRT, based on paired sample t-test. (A) All patients, (B) LOW NTCP reduction group (ΔNTCP: 0-5%), (C) HIGH NTCP reduction group (ΔNTCP >5%). Abbreviations: CI = confidence interval, IMRT = intensity modulated radiotherapy, SW-IMRT = swallowing sparing IMRT
Figure 1: NTCP comparison plot. The plot shows the NTCP values for SW-IMRT as a function of the NTCP values for standard IMRT. All dots are below the black dashed line indicating that the NTCP values for SW-IMRT are generally lower than those obtained with standard IMRT. Abbreviations: IMRT=intensity modulated radiotherapy, SW-IMRT=swallowing sparing IMRT, NTCP=normal tissue complication probability.

Model fit

In the standard IMRT and 3D-CRT patient cohort used for model development\(^7\), the overall model performance was good with a scaled Brier of 0.23 (95% CI 0.15 – 0.31) and an explained variance ($R^2$) of 0.31 (95% CI 0.21 – 0.41). Discrimination in terms of the AUC was 0.80 (95% CI 0.75 – 0.86) with a discrimination slope of 0.22 (95% CI 0.18 – 0.25). The Hosmer-Lemeshow “goodness of fit” had a p-value of 0.19, indicating a good agreement between expected and observed rates.

Similar results were seen in the validation of the current SW-IMRT patient cohort. The overall model performance had an actual scaled Brier of 0.13 (95% CI 0.03 – 0.22, with $p=0.36$ for a single sided test with respect to the expected distribution based on the model and the case-mix, such that we cannot reject the null-hypothesis that the performance is not worse than expected) and an actual explained variance ($R^2$) of 0.21 (95% CI 0.08 – 0.33, $p=0.34$). The discrimination ability of the model showed an actual AUC of 0.75 (95% CI 0.68 – 0.82, $p=0.32$) and an actual discrimination slope with a value of 0.14 (95% CI 0.10 – 0.18, $p=0.30$). The Hosmer-Lemeshow “goodness of fit” had a $p=0.74$, indicating a good agreement between expected and observed rates.
Figure 2 shows the calibration plots of the both cohorts.

**Figure 2**: Calibration plots for the previous cohort treated with standard IMRT versus that for the current cohort treated with SW-IMRT. Abbreviation: IMRT=intensity modulated radiotherapy, SW-IMRT=swallowing sparing IMRT, NTCP=normal tissue complication probability.

**Observed versus predicted NTCP values**

The mean predicted NTCP_{SW-IMRT} in all 186 patients was 22.6% (95% CI 20.2 – 24.9%), which was significantly lower than the mean predicted NTCP_{standard} which was 27.5% (95% CI 24.9 – 29.9%) (p<0.001). The observed prevalence of SWALM6 was 22.6%, which corresponded perfectly with the mean predicted NTCP values (Figure 3).

In the LOW ΔNTCP_{SWALM6} group, the mean predicted NTCP_{standard} and mean predicted NTCP_{SW-IMRT} was 23.3% (95% CI 19.2 – 27.6%) and 21.3% (95% CI 17.4 – 25.4%), respectively (p<0.001). In the LOW ΔNTCP_{SWALM6} the observed prevalence was 20.2% which corresponded with both 95%-confidence intervals of the mean predicted NTCP_{standard} and NTCP_{SW-IMRT} (Figure 3).

In the HIGH ΔNTCP_{SWALM6} group, the mean predicted NTCP_{standard} was 32.2% (95% CI 30.2 – 34.3%) while the mean predicted NTCP_{SW-IMRT} was significantly lower, i.e. 24.1% (95% CI 22.1 – 26.3%) (p<0.001). The observed prevalence of SWALM6 in the HIGH ΔNTCP_{SWALM6} was 25.3% and was within the mean predicted NTCP_{SW-IMRT} 95%-confidence interval (Figure 3).
Discussion

This study is the first to report the clinical validation of SW-IMRT using a model-based approach. Our results show that by adding dose constraints for SWOARs during treatment planning optimization a clinically relevant ΔNTCP reduction can be obtained in approximately 50% of the patients and that subsequent lower prevalence’s of SWALM6 are observed if the dose to the superior PCM and the supraglottic larynx can indeed be sufficiently decreased.

In the current study population, we were able to significantly reduce the dose the superior PCM and supraglottic larynx resulting in a significant average NTCP reduction from 27.5% obtained with standard IMRT to 22.6% as obtained with SW-IMRT. The expected average NTCP_{SW-IMRT} corresponded perfectly with the observed prevalence of 22.6%.

It should be noted that in this study, we only included patients with primary tumors originating from the oral cavity, nasopharynx, oropharynx, hypopharynx and larynx, while in our previous cohort, also other primary tumor sites were included. When confining the analysis to similar tumor sites as used in the current study, the prevalence of SWALM6 in
patients treated with standard IMRT in our previous study was 27.9%, which corresponded nicely with the mean predicted NTCP\textsubscript{standard} found in the current study (27.5%), indicating that the expected prevalence of SWAL\textsubscript{M6} would have been similar to that observed in the previous cohort when treated with standard IMRT.

The NTCP for SWAL\textsubscript{M6} with standard IMRT and the ΔNTCP\textsubscript{SWALM6} varied widely between individual patients. We found differences between the pretreatment variables of the patients in the LOW (0-5%) and HIGH (>5%) ΔNTCP\textsubscript{SWALM6} group. The HIGH ΔNTCP\textsubscript{SWALM6} group, consisted of patients with more advanced T-stages, and with more nasopharyngeal and oropharyngeal cancers, which corresponds with the higher doses administered to the most important SWOARs, i.e. the superior PCM and supraglottic larynx. Consequently, patients in the HIGH ΔNTCP\textsubscript{SWALM6} group more often received CHRT or conventional RT.

As shown in a previous report on the implementation of SW-IMRT\textsuperscript{16}, both NTCP\textsubscript{standard} and ΔNTCP obtained with SW-IMRT depend on the dose to the SWOARs obtained with standard IMRT. There are several reasons why ΔNTCP\textsubscript{SWALM6} remained low in some of the patients. In the first place, the SWOARs sometimes partly or even completely overlap with the PTV, and therefore little to no reduction in the SWOARs could be obtained without compromising the dose to the PTV. Secondly, lowering the dose to the SWOARs in some cases led to an increase of the dose to other OARs which was not allowed according to predefined criteria. Finally, for a small subgroup of patients the initial NTCP was already low and could not be reduced any further.

Others have previously reported on the relationship between dose-volume parameters and swallowing dysfunction after RT or CHRT, and on the potential benefit of sparing SWOARs with IMRT, but the clinical relevance of these reductions remained to be determined\textsuperscript{2,17–20}. Some of the other investigators who previously reported on the potential dosimetric benefits of SW-IMRT, accepted reduced coverage of the (elective) PTV (i.e. by using split field IMRT)\textsuperscript{18,19,21–23}, which makes it difficult to compare these results with those from the current study.

In the current study we used the previously described model-based approach to validate SW-IMRT. The model-based approach is a stepwise methodology, that has been developed to effectively select patients that are expected to benefit most from new radiation delivery techniques aiming at the reduction of radiation-induced side effects, such as proton therapy\textsuperscript{12}. However, this method is also applicable for the development and validation of other radiation delivery techniques, such as SW-IMRT.

At present, randomized trials are still considered gold standard in evidence-based medicine. This is certainly true for new interventions aiming at improving treatment efficacy in terms of local control and survival. However, evidence based medicine is not restricted to randomized trials and meta-analyses. It involves tracking down the best external evidence with which to answer our clinical questions\textsuperscript{24}. For new radiation techniques that are only aiming at reducing side effects without changing the strategy with regard to
tumor control (i.e. target volumes and fractionation), the model-based approach can be considered as a good alternative for an RCT.

The model-based approach consists of two phases (phase α): a phase aiming at the development of a new radiation technique (3 steps), and a consecutive phase (phase β) including a prospective observational cohort study aiming at the clinical validation of the new radiation delivery technique (step 4). For swallowing dysfunction and SW-IMRT, phase α, including multivariable NTCP model development followed by in silico planning comparisons have been reported in our previously published papers. In the current paper we reported on phase β. The multivariable NTCP model for SWALM6 showed a significant relationship between the dose to the superior PCM and supraglottic larynx and SWALM6. However, such relationship does not strictly guarantees a causal relationship between these two dose parameters and SWALM6. Therefore, an essential step in assessing the generalizability and causality of the multivariable NTCP model for SWALM6 was external validation in an independent patient cohort treated with a radiation technique in which the relevant dose parameters were further optimized. The current study showed that the multivariable NTCP model for SWALM6 still performed well in this independent patient cohort with a modified IMRT technique and that the predictions based on the actual treatment technique (SW-IMRT) corresponded significantly better with the observed prevalence than the predictions based on the standard IMRT plans.

Recently, Van der Laan et al. reported on the potential benefit of intensity-modulated proton radiotherapy (IMPT) to reduce swallowing dysfunction (SW-IMPT). In this planning comparative study the dose to the SWOARs and subsequent NTCP values for xerostomia and swallowing dysfunction could be reduced even further with SW-IMPT compared to SW-IMRT.

Conclusion

We externally validated the multivariable NTCP model for SWALM6 in a subsequent independent cohort of patients treated with SW-IMRT and showed that by reducing the dose parameters included in this NTCP model, the risk of swallowing dysfunction can be reduced.

References
