Acute symptoms during the course of head and neck radiotherapy or chemoradiation are strong predictors of late dysphagia


Department of Radiation Oncology, Department of Otorhinolaryngology/Head and Neck Surgery, Department of Medical Oncology, and Department of Oral and Maxillofacial Surgery, University of Groningen, University Medical Center Groningen, The Netherlands

Prediction of dysphagia

Acute symptoms during the course of head and neck radiotherapy or chemoradiation are strong predictors of late dysphagia. When acute symptoms during the course of RT are strong prognostic factors for late dysphagia, these symptoms impair normal swallowing, leading to weight loss and requiring dietary changes. The severity ranges from minor swallowing problems with a normal diet to complete dependence on tube feeding. Previous studies showed that dysphagia is a detrimental impact on quality of life after treatment of HNC.

Purpose: To determine if acute symptoms during definitive radiotherapy (RT) or chemoradiation (CHRT) are prognostic factors for late dysphagia in head and neck cancer (HNC).

Material and methods: This prospective cohort study consisted of 260 HNC patients who received definitive RT or CHRT. The primary endpoint was grade 2–4 swallowing dysfunction at 6 months after completion of RT (SWALM6). During treatment, acute symptoms, including oral mucositis, xerostomia and dysphagia, were scored, and the scores were accumulated weekly and entered into an existing reference model for SWALM6 that consisted of dose–volume variables only.

Results: Both acute xerostomia and dysphagia were strong prognostic factors for SWALM6. When acute scores were added as variables to the reference model, model performance increased as the course of treatment progressed: the AUC rose from 0.78 at the baseline to 0.85 in week 6. New models built for weeks 3–6 were significantly better able to identify patients with and without late dysphagia.

Conclusion: Acute xerostomia and dysphagia during the course of RT are strong prognostic factors for late dysphagia. Including accumulated acute symptom scores on a weekly basis in prediction models for late dysphagia significantly improves the identification of high-risk and low-risk patients at an early stage during treatment and might facilitate individualized treatment adaptation.

© 2015 Elsevier Ireland Ltd. All rights reserved. Radiotherapy and Oncology 115 (2015) 56–62
associated to the development of late dysphagia. The second objective was to determine if and to what extent model performance would improve as a result of including acute symptoms as prognostic factors in the reference multivariable NTCP model for SWALM6.

Materials and methods

Patients

We acquired the patient data for our study from the HNC database at our department. All patients with HNC referred for RT or CHRT are subjected to a prospective data collection program in which baseline, acute (weekly during RT) and late (six months after RT) radiation-induced side effects are assessed on a routine basis. For the current study, we included patients who received definitive RT, CHRT or RT with cetuximab for Stage I–IV (M0) squamous cell HNC. We excluded patients who had previously undergone surgery and/or RT in the head and neck region, who had prior malignancies and those with distant metastases or locoregional recurrences at 6 months after treatment. Patients with grade 1–4 dysphagia at the baseline and patients with missing data on acute symptoms in two or more subsequent weeks were also excluded. In 413 cases treated with definitive radiotherapy, chemoradiation or radiotherapy plus cetuximab, the dosimetric data and the SWALM6 scores were both available. Patients were excluded as a consequence of: (1) the exclusion criteria \( n = 82 \); (2) missing baseline toxicity data \( n = 1 \); or (3) missing acute symptom data \( n = 3 \), respectively. Eventually, of the remaining 327 patients, 67 were excluded because of a baseline dysphagia grading of more than zero. The final study population consisted of 260 patients in whom all pretreatment, treatment and toxicity data were prospectively assessed. Table 1 shows the demographic, tumor and treatment characteristics of the study population.

All patient data was obtained as part of a prospective data registration program within the framework of routine clinical practice. The Dutch Medical Research Involving Human Subjects Act is not applicable to data collection as part of routine clinical practice. Therefore the hospital ethics committee exempted this study from the ethical approval requirement.

Endpoint and acute symptoms

The primary endpoint was physician-rated dysphagia grade 2–4 at 6 months after the completion of RT (SWALM6) according to the Late European Organisation for Research and Treatment of Cancer/ Radiation Therapy Oncology Group (EORTC/RT0G) Radiation Morbidity Scoring Criteria [9]. Patients with this endpoint were unable to eat solid food and could only eat semisolid (pureed) food (grade 2) or worse. Acute and late radiation-induced side effects were assessed at the baseline and weekly during the course of treatment. The acute symptom scores for oral mucositis, xerostomia and dysphagia are listed in Table 2. If acute symptom scores were missing for only one week for a particular patient, the scores from the previous week were used.

Treatment

Patients were treated with three-dimensional conformal radiotherapy (3D-CRT) or intensity modulated radiotherapy (IMRT), with or without concomitant chemotherapy or cetuximab. These treatment regimes have been described previously in more detail [6]. The various organs at risk were delineated on the planning CT according to previously published delineation guidelines [10,11]. All original 3D-CRT and IMRT treatment plans, with the corresponding dose-volume parameters of the organs at risk, were evaluated in a research version of the Pinnacle\(^3\) treatment-planning system (version 9.1, Philips Radiation Oncology Systems, Fitchburg, WI, USA).

Reference model

First, we evaluated the reference multivariable NTCP model in our cohort (Table 1) by calculating the model value (linear predictor) for each patient with the model intercept and regression coefficients as published by Christianen et al. These values were \(-6.09 + (0.057 \times \text{mean dose (Gy)} \) in the superior PCM) + (0.037 \times \text{mean dose (Gy)} \) in the supraglottic larynx), and the subsequent NTCP value for each individual patient. In addition, we evaluated the performance of this model in the current study population relative to that in Christianen et al. [6] in terms of explained variance, calibration and discrimination. The linear predictor of the reference model was entered as the reference (baseline) model variable in all subsequent models.

Standardized acute symptoms z-scores

The acute symptom scores were assessed weekly (Table 2). Each acute symptom score was subsequently transformed into a weekly standardized accumulated z-score for two reasons: (1) an accumulated score, calculated by adding the scores of previous weeks, combined the severity and duration of the acute symptom in one value;
standardizing the accumulated scores (i.e., expressing each accumulated score as a number of standard deviations from the average population score) improved the interpretation of a score.

Formulas

For each of the acute symptoms and individual patient \( k \), an accumulated symptom score \( C_{k,w} \) in a particular week of radiotherapy \( w \) was calculated by adding up the scores \( S_{kj} \) of all subsequent weeks \( j \) from week 1 to week \( w \).

\[
C_{k,w} = \sum_{j=1}^{w} S_{kj}
\]

For each week of radiotherapy \( w \), the mean accumulated acute symptom score \( \mu_w \) and standard deviation \( \sigma_w \) were calculated for the whole population of \( n \) patients.

\[
\mu_w = \frac{1}{n} \sum_{k=1}^{n} C_{k,w}
\]

\[
\sigma_w = \sqrt{\frac{\sum_{k=1}^{n} (C_{k,w} - \mu_w)^2}{n-1}}
\]

Finally, a standardized z-score \( Z_{k,w} \) for the accumulated symptom scores of patient \( k \) in week \( w \) was defined as

\[
Z_{k,w} = (C_{k,w} - \mu_w)/\sigma_w
\]

Weekly dynamic models

The acute symptom profile of individual patients became clearer as RT progressed. Because new acute symptom information was added each week, these weekly models were referred to as dynamic models.

The z-scores were available for each patient, each week of RT and each acute symptom type (oral mucositis, xerostomia and dysphagia). The next step was to build subsequent dynamic models for each week (weeks 1, 2, 3, 4, 5, and 6). The variables in a dynamic model consisted of the aforementioned linear predictor of the reference model and the z-scores for the acute symptoms. Then, a stepwise backward (Wald) logistic regression procedure was used to exclude acute symptom variables from the model with \( p \)-removal >0.157.

Internal validation and model performance

All dynamic models were subjected to internal validation with a bootstrapping procedure (2000 bootstraps for each analysis) in order to correct (shrink) the models (slope and intercept) for optimism [12]. This was done to obtain realistic regression coefficients for the model variables that are representative for populations similar to the development sample. For each dynamic model, the area under the receiver operating characteristic curve (ROC-curve AUC), the explained variance and calibration were determined. In addition, model improvement measures were determined, i.e., the Net Reclassification Improvement (NRI) and the Integrated Discrimination Improvement (IDI) relative to the reference model were determined for each dynamic model. The NRI quantified the difference in the sum of the sensitivity and specificity between two models (using a >50% NTCP criterion to classify patients as high-risk patients). The IDI quantified the difference in the discrimination slopes of two models. Dynamic models were only accepted for a particular week when these yielded a significantly better \( (p < 0.05) \) classification and discrimination (NRI and IDI) of patients with and without late dysphagia on the basis of the model predictions compared with the existing reference model.

NTCP calculation

The probability that a patient developed the complication (NTCP), i.e., late dysphagia, was calculated with a logistic regression model
\[ \text{NTCP} = \frac{1}{1 + e^{-S}} \]

with the linear predictor \( S \) defined as
\[ S = \beta_0 + \sum \beta_i x_i \]

where \( \beta_0 \) and \( \beta_i \) were the model parameters and \( x_i \) the predictor variables.

**Results**

**Reference model**

Of the 260 patients included in the final analyses (Table 1), 63 (24.2%) developed SWALM6 (Table 2). The linear predictor variable obtained from the reference model of Christianen et al. (comprising the intercept, the mean dose in the superior PCM and the mean dose in the supraglottic larynx with corresponding regression coefficients) was a significant factor in all subsequent models in the current study (Table 3). The reference model yielded an AUC value of 0.78 in our cohort (Table 4). This value was comparable (0.80) to the AUC obtained by Christianen et al. [6].

**Discarded models and predictors**

In univariable analysis, all acute symptom \( z \)-scores in weeks 2 through 6 (and acute xerostomia in weeks 1 through 6) were significantly associated with SWALM6 (Table 3). The linear predictor variable (coefficients) was a significant factor in all subsequent models in the multivariable models, acute xerostomia in weeks 1 and 2 was a significant factor for SWALM6. Although the performance of the week 1 and 2 models improved compared to the reference model in our cohort (ROC-curve AUC: 0.792 and 0.804, \( R^2 \): 0.276 and 0.300, respectively), these models did not result in a significant NRI, nor a significant IDI compared to the reference model.

**Dynamic models**

Acute xerostomia and acute dysphagia in weeks 3, 4, 5 and 6 were significant prognostic factors for SWALM6 (Table 3). The dynamic NTCP models in these weeks all performed better than the reference NTCP model in our cohort. Classification of high-risk patients was significantly better with the dynamic models (Table 4). For example, the NRI ranged from 0.212 in week 3 to 0.303 in week 6 (\( p < 0.001 \)). The discrimination slope of the reference model was 0.165, while in the dynamic models it ranged from 0.247 in week 3, to 0.301 in week 6. This corresponds with a statistically significant IDI (\( p < 0.001 \)).

The duration and severity of acute xerostomia and acute dysphagia strongly affected the NTCP values of SWALM6 (Fig. 1). For example, in the quartile of patients who had the worst acute symptoms in week 4, SWALM6 was observed in 54% of the cases. The week 4 model predicted an average NTCP of 54% in this group, compared to 34% on average with the reference model. In the quartile of patients who had the mildest acute symptoms in week 4, SWALM6 was observed in 3% of the cases. The week 4 model predicted an average NTCP of 5% in this group, and the reference model 13% on average.

We verified if other variables, listed in Table 1, improved the final dynamic models as presented in this paper, which however was not the case. This finding corresponds with the previous findings during the development of the reference model.

**Table 3**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Uncorrected</th>
<th>Corrected</th>
<th>OR</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reference model in current cohort</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear predictor reference model</td>
<td>0.881</td>
<td>0.891</td>
<td>2.412</td>
<td>1.734–3.355</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Intercept</td>
<td>−0.113</td>
<td>−0.105</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model week 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear predictor reference model</td>
<td>0.739</td>
<td>0.692</td>
<td>2.094</td>
<td>1.483–2.955</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Xerostomia ( z )-value</td>
<td>0.427</td>
<td>0.400</td>
<td>1.533</td>
<td>1.047–2.243</td>
<td>0.028</td>
</tr>
<tr>
<td>Dysphagia ( z )-value</td>
<td>0.483</td>
<td>0.452</td>
<td>1.620</td>
<td>1.148–2.288</td>
<td>0.006</td>
</tr>
<tr>
<td>Intercept</td>
<td>−0.419</td>
<td>−0.444</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model week 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear predictor reference model</td>
<td>0.655</td>
<td>0.626</td>
<td>1.926</td>
<td>1.354–2.739</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Xerostomia ( z )-value</td>
<td>0.368</td>
<td>0.351</td>
<td>1.445</td>
<td>0.945–2.209</td>
<td>0.089</td>
</tr>
<tr>
<td>Dysphagia ( z )-value</td>
<td>0.741</td>
<td>0.708</td>
<td>2.099</td>
<td>1.426–3.089</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Intercept</td>
<td>−0.614</td>
<td>−0.622</td>
<td>0.541</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model week 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear predictor reference model</td>
<td>0.612</td>
<td>0.587</td>
<td>1.844</td>
<td>1.293–2.629</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Xerostomia ( z )-value</td>
<td>0.458</td>
<td>0.439</td>
<td>1.581</td>
<td>1.004–2.491</td>
<td>0.048</td>
</tr>
<tr>
<td>Dysphagia ( z )-value</td>
<td>0.725</td>
<td>0.696</td>
<td>2.066</td>
<td>1.383–3.086</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Intercept</td>
<td>−0.736</td>
<td>−0.737</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model week 6</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear predictor reference model</td>
<td>0.556</td>
<td>0.531</td>
<td>1.743</td>
<td>1.213–2.505</td>
<td>0.003</td>
</tr>
<tr>
<td>Xerostomia ( z )-value</td>
<td>0.559</td>
<td>0.534</td>
<td>1.749</td>
<td>1.077–2.840</td>
<td>0.024</td>
</tr>
<tr>
<td>Dysphagia ( z )-value</td>
<td>0.768</td>
<td>0.734</td>
<td>2.155</td>
<td>1.410–3.295</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Intercept</td>
<td>−0.885</td>
<td>−0.878</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Binary endpoint: physician-rated dysphagia grade 2–4 at 6 months after the completion of radiotherapy according to the Late EORTC/RTOG Radiation Morbidity Scoring Criteria [9].

**Abbreviations:** \( \beta \) = regression coefficient; EORTC/RTOG = European Organisation for Research and Treatment of Cancer / Radiation Therapy Oncology Group. OR = Odds Ratio; \( z \)-value = Standardized mean difference of the accumulated toxicity score (\( (\text{accumulated patient score} - \text{group mean accumulated score})/\text{group standard deviation accumulated score} \)).

* In addition to the uncorrected model parameters the corrected \( \beta \) values after internal validation and model shrinkage are shown.
Discussion

The main objective of the current study was to test the hypothesis that acute symptoms during the course of treatment are significantly associated to the development of late dysphagia. This hypothesis was supported by our data. Acute dysphagia and acute xerostomia in weeks 3–6 of radiotherapy were independent prognostic factors for late RTOG grade 2–4 dysphagia. The second objective concerned the performance of the reference multivariable NTCP model for SWALM6. This performance improved significantly when the weekly accumulated scores for acute dysphagia and acute xerostomia were added as variables: the dynamic models were significantly better at distinguishing patients who developed late dysphagia from those who did not.

These dynamic models could be invaluable in the context of a more personalized treatment approach. Current dose schedules are often based on the average toxicity observed in the total patient population, while our data indicate that, even with the same dose in the SWOARs, some patients sustain more acute dysphagia than average. We demonstrated that these patients also have a higher risk of late dysphagia, and might therefore be considered as candidates for a mid-course re-evaluation of their treatment regimes (dose prescriptions and/or radiotherapy techniques). Several potential means are available to do so. For example, in week 4, when patients develop more than average acute symptoms, the dose in the swallowing structures could be lowered for the remaining fractions, e.g., by decreasing the total dose in the elective target volumes [13], by using adaptive

### Table 4

<table>
<thead>
<tr>
<th>Performance measure</th>
<th>Reference model</th>
<th>Week 3 model</th>
<th>Week 4 model</th>
<th>Week 5 model</th>
<th>Week 6 model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall –2 log likelihood</td>
<td>239.1</td>
<td>219.6</td>
<td>209.6</td>
<td>209.4</td>
<td>206.3</td>
</tr>
<tr>
<td>Nagelkerke adjusted R²</td>
<td>0.256</td>
<td>0.345</td>
<td>0.388</td>
<td>0.389</td>
<td>0.402</td>
</tr>
<tr>
<td>Discrimination</td>
<td>0.781 (0.721–0.840)</td>
<td>0.823 (0.768–0.878)</td>
<td>0.847 (0.796–0.898)</td>
<td>0.846 (0.794–0.897)</td>
<td>0.849 (0.797–0.901)</td>
</tr>
<tr>
<td>ROC-curve AUC (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calibration Disc. slope</td>
<td>0.165</td>
<td>0.247</td>
<td>0.288</td>
<td>0.289</td>
<td>0.301</td>
</tr>
<tr>
<td>Hosmer–Lemeshow test</td>
<td>(p = 0.250)</td>
<td>(p = 0.775)</td>
<td>(p = 0.503)</td>
<td>(p = 0.945)</td>
<td>(p = 0.728)</td>
</tr>
<tr>
<td>IDI</td>
<td></td>
<td>0.082 (p &lt; 0.001)</td>
<td>0.122 (p &lt; 0.001)</td>
<td>0.124 (p &lt; 0.001)</td>
<td>0.135 (p &lt; 0.001)</td>
</tr>
<tr>
<td>NRI index</td>
<td></td>
<td>0.212 (p = 0.006)</td>
<td>0.250 (p = 0.002)</td>
<td>0.281 (p &lt; 0.001)</td>
<td>0.303 (p &lt; 0.001)</td>
</tr>
</tbody>
</table>

Abbreviations: IDI = Integrated Discrimination Improvement (discrimination slope dynamic model — discrimination slope baseline model); NRI = Net Reclassification Improvement (compared to the baseline model and classifying patients with a predicted risk >50% as high risk patients); ROC-curve AUC = area under the receiver operating characteristics curve.
replanning or by changing to other treatment modalities such as proton therapy (where available) [14]. Conversely, patients who remain free from acute symptoms during the entire course of treatment might be candidates for dose escalation strategies. As the safety and efficacy of such treatment adjustments are not yet clear, these kind of tailored treatment adjustments based on acute symptom profiles may be worthwhile to investigate in future clinical studies.

The rationale behind the main hypothesis of this study was that patients differ in their sensitivity to radiation treatment [15,16], and that patients who are more susceptible to radiation-induced effects are more prone to develop both acute symptoms and late toxicity. Acute symptoms can therefore be regarded as biomarkers for late toxicities and can be used to select patients for adjustment of their treatment parameters, as described above.

Another possible explanation for the findings in our study is consequential late toxicity [17–20]. This is based on the assumption that the radiation dose causes both acute and late toxicity, while the acute toxicity by itself sets in motion a series of effects that ultimately lead to late toxicity. This implies that acute toxicities are at least partly in the causal pathway from radiation dose to late toxicity. Such causality cannot be proven from observational data, and should therefore be studied and established in a biological experimental setting. However, the regression coefficient of the reference model variable became considerably smaller when the acute symptom variables were added to the model. This reduction ranged from 16% in week 3–37% in week 6. This agrees with the possibility that the observed late dysphagia is partially a consequential effect, i.e., caused by acute dysphagia [21,22]. Moreover, it is not entirely clear that the same genetic alterations responsible for radiation sensitivity result in increased acute and late toxicity. It may very well be that some alterations result in sensitivity to acute effects and some to late effects. Our data seem to support that these may be linked. The subset of patients with severe acute toxicity could be the subject of future research with the aim to find out what the biologic cause of their inherit sensitivity is. Another explanation for the variation in observed acute symptoms and late dysphagia might be that the actual delivered dose distribution was not always similar to the planned dose distribution, e.g., changes in patient anatomy may occur during the course of treatment. We are currently investigating such issues as part of an ongoing effort to improve our models.

The linear predictor of the reference model [6], including baseline dose–volume parameters, was used as a variable in all our models. We decided to take this approach instead of fitting a new baseline model to our data to ensure that the outcomes of this study added to the reference model for SWALM6. We verified that this choice did not have an important effect on the results and conclusions of the current study. All models would have fitted slightly better and would have yielded slightly better performance measures, while the differences between the models would remain similar.

Other authors have shown that baseline dysphagia is also a strong prognostic factor for late dysphagia [5,23]. In most cases, dysphagia before the start of RT is caused by other factors, such as local tumor extension or surgery. As we intended to focus on radiation-induced dysphagia, patients with grade 1–4 dysphagia at the baseline were excluded from the analysis. This also accounts for the relatively low incidence (24.2%) of late dysphagia in the current study. Therefore, it should be noted that the models presented in the current study only apply to patients who receive definitive RT for squamous cell head and neck cancer and did not have grade 1–4 dysphagia at the baseline.

In the current study, we used cumulative z-scores as candidate variables for the dynamic prediction models. In patients with other anatomical tumor sites it has been shown that accumulated acute symptom scores are more predictive of late symptoms than peak acute score changes [24]. A cumulative score also offers the ability to express the onset, severity and persistency of the acute symptoms in a single value. The disadvantage of such a value is that it may be hard to interpret and may be different for different toxicities and different scoring systems. To account for that, the acute symptom scores were accumulated each week and then transformed into standardized z-scores [25], and subsequently entered as variables in the weekly models.

Conclusion

Accumulated acute symptoms during the course of RT appear to be strong predictors of late dysphagia. With the proposed dynamic models, the NTCP of late dysphagia can be estimated and adjusted on a weekly basis during the course of treatment. For individual patients, this may reveal that they are more prone to late dysphagia than expected on the basis of a reference (baseline factors) model. Or conversely, when they develop no acute symptoms at all, they may be less prone to late dysphagia. The dynamic models provide valuable insight into individual therapeutic windows. This new information may potentially be used to adjust the treatment strategy mid-course on the basis of the expected risk of late dysphagia, enabling a more personalized treatment approach.

Conflict of interest statement

The authors state that the research presented in this manuscript is free of conflicts of interest.

References

Acute symptoms predict late dysphagia


