Clinical Oncology xxx (2016) 1–11

Contents lists available at ScienceDirect

Clinical Oncology

journal homepage: www.clinicaloncologyonline.net

Original Article

Normal Tissue Complication Probability (NTCP) Modelling of Severe Acute Mucositis using a Novel Oral Mucosal Surface Organ at Risk

J.A. Dean *, L.C. Welsh †, K.H. Wong †, A. Aleksic †, E. Dunne †, M.R. Islam †, A. Patel †, P. Patel †, I. Petkar †, I. Phillips †, J. Sham †, U. Schick †, K.L. Newbold †‡, S.A. Bhide †‡, K.J. Harrington †‡, C.M. Nutting †‡, S.L. Gulliford *

* *Joint Department of Physics at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, London, UK*
† *Head and Neck Unit, The Royal Marsden NHS Foundation Trust, London, UK*
‡ *Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, UK*

Received 12 July 2016; received in revised form 20 October 2016; accepted 1 November 2016

Abstract

Aims: A normal tissue complication probability (NTCP) model of severe acute mucositis would be highly useful to guide clinical decision making and inform radiotherapy planning. We aimed to improve upon our previous model by using a novel oral mucosal surface organ at risk (OAR) in place of an oral cavity OAR.
Materials and methods: Predictive models of severe acute mucositis were generated using radiotherapy dose to the oral cavity OAR or mucosal surface OAR and clinical data. Penalised logistic regression and random forest classification (RFC) models were generated for both OARs and compared. Internal validation was carried out with 100-iteration stratified shuffle split cross-validation, using multiple metrics to assess different aspects of model performance. Associations between treatment covariates and severe mucositis were explored using RFC feature importance.
Results: Penalised logistic regression and RFC models using the oral cavity OAR performed at least as well as the models using mucosal surface OAR. Associations between dose metrics and severe mucositis were similar between the mucosal surface and oral cavity models. The volumes of oral cavity or mucosal surface receiving intermediate and high doses were most strongly associated with severe mucositis.
Conclusions: The simpler oral cavity OAR should be preferred over the mucosal surface OAR for NTCP modelling of severe mucositis. We recommend minimising the volume of mucosa receiving intermediate and high doses, where possible.

Key words: Head and neck radiotherapy; machine learning; mucositis; NTCP modelling; OAR delineation; oral mucosa

Introduction

Mucositis is a common and important acute toxicity of head and neck radiotherapy, which may result in pain, dysphagia [1] and weight loss, and, hence, reduced quality of life [2,3]. Mucositis may lead to missed treatment fractions [4] and is frequently dose limiting in dose-escalation and accelerated fractionation regimens designed to improve tumour control [5–7]. Furthermore, severe acute reactions have been implicated in the subsequent development of 'late' radiation toxicity [8–10]. A normal tissue complication probability (NTCP) model for severe mucositis, with sufficient predictive performance, could be used for clinical decision-support [11]. Associations between radiotherapy dose metrics and mucositis could inform changes to the radiotherapy planning dose objectives to reduce the incidence of severe mucositis. It has previously been shown that intensity-modulated radiotherapy can be used to spare the oral mucosa in oropharyngeal radiotherapy patients [12].

Our group has previously generated and internally validated an acute mucositis NTCP model, with modest-to-good discriminative ability (mean area under the receiver operating characteristic curve [AUC] = 0.71, standard deviation = 0.09 on internal validation) [13]. The below-perfect predictive performance was attributed to limitations of the organ at risk (OAR) segmentation, limitations of the toxicity scoring instrument and not having data on

Author for correspondence: J.A. Dean, Joint Department of Physics at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, London SM2 5NG, UK. Tel: +44-208-9156223; Fax: +44-208-6433812.
E-mail address: jamie.dean@icr.ac.uk (J.A. Dean).

http://dx.doi.org/10.1016/j.clon.2016.12.001

relevant clinical and biological parameters. Only the first of these can be addressed retrospectively. We hypothesised that the discriminative ability could be improved through a more accurate description of the dose distribution to the oral mucosal surfaces, which were previously described by the oral cavity volume contours (OCC) method [14]. The OCC OAR predominantly describes the dose to the musculature of the tongue and floor of mouth, which are not expected to be relevant for mucositis, and does not incorporate the buccal mucosa or mucosa of the lips. We, therefore, developed a novel oral mucosal surface contours (MSC) OAR structure that includes the mucosal surfaces of the oral cavity and excludes the musculature of the tongue and floor of mouth [15]. It was expected that this novel OAR would be more relevant to NTCP modelling of mucositis. To aid implementation of this OAR to a large cohort of patients, we showed that the segmentation of this structure could be fully automated using atlas-based segmentation [16].

The aims of this study were to: (i) improve the discriminative ability of our, previously reported, acute mucositis NTCP model [13] and (ii) gain greater insight into the radiotherapy dose–response relationship of the oral mucosa, through the application of our novel, automatically segmented MSC structure [15,16].

Materials and Methods

Patient Data

A cohort of 351 head and neck radiotherapy patients, enrolled in six different clinical trials [17–23] (with institutional review board approval and signed patient consent; summarised in Appendix A), diverse in terms of primary disease site, radiotherapy treatment technique and use of concurrent chemotherapy, was used. The prescribed dose to the primary planning target volume was either 65 Gy in 30 fractions, 60 Gy in 30 fractions, 67.2 Gy in 28 fractions or 63 Gy in 28 fractions (described in Appendix A). Toxicity was consistently scored for all studies using the mucositis score from the Common Terminology Criteria for Adverse Events (CTCAE) versions 2 (mucositis due to radiation) [24] or 3 (mucositis/stomatitis [clinical examination]) [25] instruments, which are near equivalent. Toxicities were recorded prospectively before the start of radiotherapy, weekly during radiotherapy, and at 1–4 and 8 weeks after radiotherapy. The toxicity end point of interest chosen for analysis was the maximum reported mucositis grade. Patients were dichotomised into severe (maximum toxicity score of grade 3 or worse) and non-severe (maximum toxicity score of less than grade 3) mucositis. Patients with baseline toxicity or any missing toxicity scores and a maximum score below 3 were excluded from the analysis. Our missing data handling strategy is discussed in Appendix B. Complete DICOM radiotherapy data were available for 351 patients. After removing patients who had both missing toxicity data and a maximum toxicity of grade 2 or lower, 182 patients were available for analysis. MSC atlas-based segmentation failed for three patients (for no obvious reasons), so these were excluded from the analysis, leaving 179. Severe mucositis incidence was 74%. Note that excluding the patients removed due to missing toxicity data skews the incidence values to higher than the actual incidences.

Induction chemotherapy ($n = 89$), concurrent chemotherapy regimen (cisplatin [$n = 64$], carboplatin [$n = 10$], one cycle of cisplatin followed by one cycle of carboplatin [$n = 6$] or none [$n = 99$]; administered in two cycles, on days 1 and 29 of radiotherapy), definitive ($n = 149$) versus postoperative radiotherapy, age (median = 57 years, range = 17–88 years), gender ($n_{male} = 114$) and primary disease site (nasopharynx [$n = 18$], oropharynx [$n = 100$], hypopharynx/larynx [$n = 18$], unknown primary [$n = 8$] or parotid gland [$n = 35$]) were also included as covariates in the models. Unilateral versus bilateral irradiation was not explicitly included as a covariate in the models as it correlates perfectly with parotid gland primary disease site.

Radiotherapy Dose Data

The oral mucosa was contoured on computed tomography using two different techniques: the current guidelines (OCC) method [14] and our novel MSC technique [15,16]. Mucositis of the portion of the pharyngeal mucosa visible on clinical examination was included in the scoring of mucositis. Therefore, the pharyngeal mucosa was manually delineated from the roof of the nasopharynx to the level of the inferior border of the oral mucosa structure (OCC or MSC as appropriate) and combined with the OCC or MSC structure (denoted OCC-PM and MSC-PM). The inferior extents of the OCC and MSC structures were very similar. Figure 1 shows an example of these structures. Clinical oncologists carried out the OCC and PM contouring, following the same guidelines, using the RayStation research version 4.6.100.12 treatment planning system (RaySearch Laboratories AB, Stockholm, Sweden). MSC contouring was carried out manually for 41 patients (those included in the atlas and test cohort in [16]). For the remaining patients the MSC was delineated fully automatically. The automatically generated OARs were visually assessed for gross errors. Three patients were excluded due to failure of the automatic segmentation, as previously described in the patient data section. The techniques for manual and automatic MSC segmentation are described in [16].

The physical dose distribution was converted to the fractional dose distribution (physical dose delivered in each fraction), as recommended for modelling of acute toxicity by Tucker *et al.* [26]. The fractional dose distribution was described by the dose-volume histogram (DVH) in 20 cGy intervals from 20 to 260 cGy per fraction. There may be regional variations in radiosensitivity across the oral mucosa that cannot be detected by describing the dose distributions using only DVHs. For example, keratinised areas of the oral mucosa might be expected to be associated with lower mucositis scores [27] than non-keratinised regions of the oral mucosa [28]. Therefore, three-dimensional

Please cite this article in press as: Dean JA, et al., Normal Tissue Complication Probability (NTCP) Modelling of Severe Acute Mucositis using a Novel Oral Mucosal Surface Organ at Risk, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.12.001

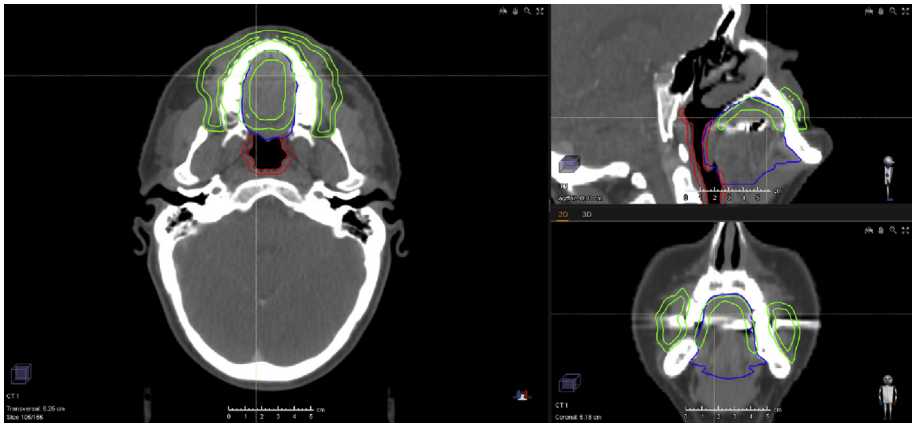

Fig 1. Example of oral cavity (OCC; blue), mucosal surface (MSC; green) and pharyngeal mucosa (PM; red) contours.

moment invariants, η_{abc} [13,29] describing the spatial distribution of the dose were calculated using the expression

$$\eta_{abc} = \frac{\mu_{abc}}{\mu_{000}^{\frac{a+b+c}{3}+1}} \quad (1)$$

where

$$\mu_{abc} = \sum_x \sum_y \sum_z |(x-\overline{x})|^a (y-\overline{y})^b (z-\overline{z})^c D(x,y,z) I(x,y,z) \quad (2)$$

where x, y and z are the voxel coordinates, $D(x,y,z)$ is the dose delivered to the voxel with coordinates (x,y,z), $I(x,y,z)$ is an identity function, which takes a value of 1 if the voxel belongs to the OAR and 0 if it does not, and $(\overline{x},\overline{y},\overline{z})$ is the centre of gravity of the OAR. The moments are translational and scale invariant. The left–right symmetry is accounted for by taking the modulus of the $(x-\overline{x})$ term. Therefore, the moments in the left–right direction describe how lateralised or centralised the dose is. The different three-dimensional moment invariants describe the centre of mass (η_{001}, η_{010}, η_{100}, η_{011}, η_{101}, η_{110}, η_{111}), spread (η_{002}, η_{020}, η_{200}) and skewness (η_{003}, η_{030}, η_{003}) of the dose distribution in the three orthogonal directions (left–right, anterior–posterior, superior–inferior) within each structure. These allow for regional variations in radiosensitivity to be probed. These would manifest as differences in one or more of the moment invariants between patients who experienced severe mucositis and those who did not. An example of the three-dimensional moment invariants is shown in Appendix C. The dose metrics were used as covariates in the statistical modelling.

Statistical Analysis

All radiotherapy dose and clinical covariates were transformed to standardised scores (mean = 0, standard deviation = 1) to avoid scale-related feature dominance. To determine which of the structures resulted in the highest discriminative ability, penalised logistic regression (PLR) [30,31] models were generated and internally validated for both of the structures (OCC-PM and MSC-PM). PLR models reduce the regression coefficients to reduce the risk of overfitting the data. Models were generated with (spa) and without (sta) the addition of the spatial dose metrics. During model generation, the samples of each outcome class were over- or undersampled according to weights inversely proportional to the class frequencies in the training data to account for unbalanced numbers of patients experiencing severe and non-severe mucositis. Model hyper-parameter tuning was carried out using a cross-validated grid-search with 100 iterations of stratified shuffle split cross-validation with a train/test split of 80/20. The possible hyper-parameters over which the cross-validated grid-searchers were carried out are given in Appendix D. To establish whether the discriminative ability could be improved through the use of a more complex model, random forest classification (RFC) [32] models were also generated (our choice of RFC models is discussed in [13]).

The generalisability of the models (aim i) was measured through internal validation. Randomly splitting datasets into a development and independent validation sample, although often carried out, provides little additional information over internal validation and increases the risk of bias in small datasets [33,34]. Therefore, all available data were used for generating and internally validating the models. External validation of any promising models should be carried out in the future. Internal validation used a nested 100-iteration stratified shuffle split cross-validation, with a train/test split of 80/20, incorporating covariate transformation to standardised scores and hyper-parameter tuning with a five-fold cross-validated grid-search within each iteration to give unbiased error estimates. AUC was used as the scoring metric for the grid-search. The predictive performance of the models was assessed using AUC to measure discrimination, Brier score [35] to measure overall model performance and log loss [36] to assess the model probability estimates. The slope and intercept of a logistic regression model of the actual mucositis outcomes against the predicted probabilities of severe mucositis made by the models was used to measure

Please cite this article in press as: Dean JA, et al., Normal Tissue Complication Probability (NTCP) Modelling of Severe Acute Mucositis using a Novel Oral Mucosal Surface Organ at Risk, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.12.001

calibration [37,38]. Consideration of discrimination and calibration metrics and their 95% confidence intervals were used for model comparisons. Selecting models based on formal statistical testing of differences in AUC is inappropriate in this context as this metric gives equal importance weighting to sensitivity and specificity, which does not align with clinical objectives.

To determine associations between treatment factors and mucositis severity (aim ii), the $RFC_{OCC\text{-}PM,spa}$ and $RFC_{MSC\text{-}PM,spa}$ model Gini feature importance values were bootstrapped with 2000 replicates. We have previously shown that this approach is more suitable than the, often used, logistic regression coefficients, in the context of highly correlated radiotherapy dose-volume metrics [13]. The statistical analysis was carried out using the Python version 2.7.9 programming language [39] and Pandas version 0.18 [40] and Scikit-learn version 0.17 [41] modules.

Results

The DVH data and correlation matrix for the data are summarised in Figure 2 and Appendix E, respectively, and show a relationship between the DVH and mucositis severity for both OARs. The correlation matrix indicates that the DVH-based covariates were highly correlated, both within the same OAR and between the two different OARs. Addressing the first aim of generating models to make accurate predictions of mucositis severity, the metrics describing the predictive performance of the models are shown in Table 1. The discriminative abilities of all of the models were modest-to-good. For all models the use of the MSC-PM OAR did not lead to an improvement in predictive performance compared with the corresponding model using the OCC-PM OAR. For both structures (OCC-PM and MSC-PM) and both types of model (PLR and RFC), the addition of the spatial dose metrics did not result in improved model performance, as assessed by any of the metrics. The RFC models were better-calibrated (calibration slope closer to 1 and calibration intercept closer to 0) and provided better probability estimates (lower log loss) and overall performance (lower Brier score) than the PLR models. However, they had slightly worse discrimination (lower AUC). Preliminary analysis indicated that excluding the pharyngeal mucosa from the OARs did not substantially alter any of the results.

The regression coefficients, means and standard deviations for the covariates in the $PLR_{OCC\text{-}PM,sta}$ and $PLR_{MSC\text{-}PM,sta}$ models are given in Table 2. These values are required to use the models. It should be noted that the regression coefficients of PLR models, like other linear models, are unstable in the presence of multicollinearity. This does not prevent them from being used to make accurate predictions of patient toxicity outcomes, but does mean that the regression coefficients should not be used to infer associations between the correlated dose metrics and severe mucositis. RFC models were used for this purpose as they are more robust to multicollinearity.

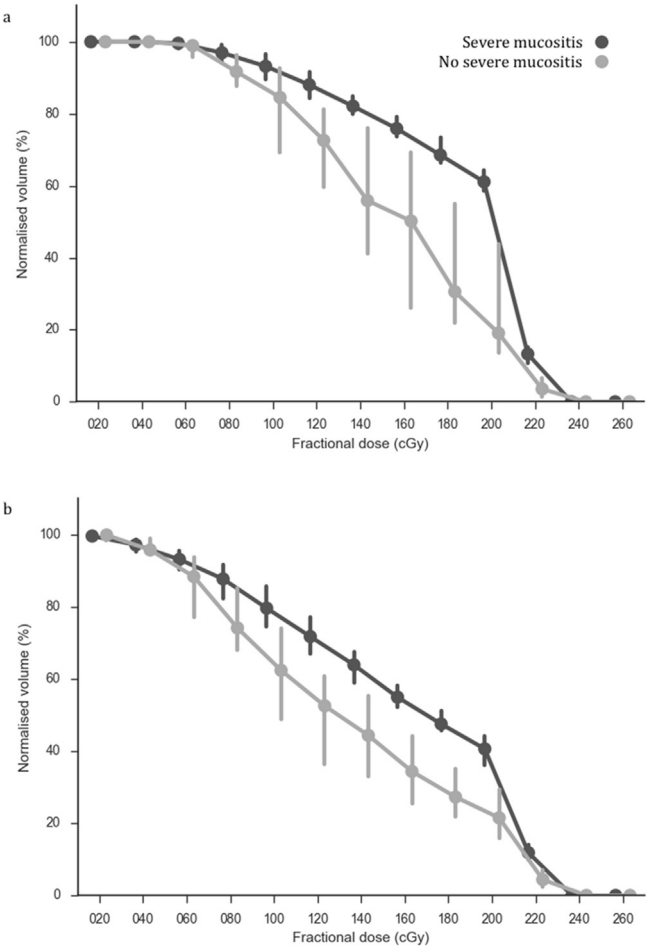

Fig 2. Summary of dose-volume data for (a) the oral cavity contours and pharyngeal mucosa organ at risk (OCC-PM) structure and (b) the mucosal surface contours and pharyngeal mucosa organ at risk (MSC-PM) structure, grouped by peak mucositis severity. The lines represent the group medians and the error bars represent the bootstrapped 95 percentile confidence intervals on the medians. It should be noted that the volumes are measured at the same fractional dose values for both outcome groups. However, there is an artificial offset to aid visualisation of overlapping error bars.

Regarding the second aim of establishing associations between treatment covariates, such as radiotherapy dose metrics, and mucositis severity, the feature importance values of the RFC models are displayed in Figure 3. These indicate that, when the OCC-PM was considered as the OAR, the feature importance of the dose-volume metrics increased with increasing dose, peaking at V180 and V220. The spatial dose metric with the highest feature importance was η_{010}, which represents the centre of mass of the dose in the anterior–posterior direction. When the MSC-PM was considered as the OAR, V160–V220 had the highest feature importance values. A sharp increase in feature importance was observed between V120 and V160. Feature importance declined from V220 to V260. The decline in feature importance does not indicate a reduction in biological effect at these dose levels, rather it is due to a lack of variance in

Please cite this article in press as: Dean JA, et al., Normal Tissue Complication Probability (NTCP) Modelling of Severe Acute Mucositis using a Novel Oral Mucosal Surface Organ at Risk, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.12.001

Table 1
Measures of predictive performance of models on internal validation

Model	Hyper- parameters	Mean (standard deviation)				
		AUC	Log loss	Brier score	Calibration slope	Calibration intercept
$PLR_{OCC\text{-}PM,sta}$	Penalty = l1; C = 0.1	0.71 (0.10)	0.66 (0.04)	0.23 (0.02)	11.3 (11.2)	−4.4 (5.5)
$PLR_{MSC\text{-}PM,sta}$	Penalty = l2; C = 0.001	0.70 (0.09)	0.66 (0.04)	0.23 (0.01)	14.5 (12.1)	−6.0 (5.9)
$PLR_{OCC\text{-}PM,spa}$	Penalty = l1; C = 0.1	0.71 (0.11)	0.67 (0.05)	0.23 (0.02)	10.7 (10.9)	−4.1 (5.4)
$PLR_{MSC\text{-}PM,spa}$	Penalty = l1; C = 0.1	0.69 (0.09)	0.67 (0.05)	0.23 (0.02)	11.1 (10.3)	−4.3 (5.0)
$RFC_{OCC\text{-}PM,sta}$	Maximum depth = 5, maximum features = square root	0.69 (0.09)	0.56 (0.08)	0.19 (0.03)	3.3 (2.1)	−1.0 (1.3)
$RFC_{MSC\text{-}PM,sta}$	Maximum depth = 5, maximum features = square root	0.68 (0.08)	0.56 (0.06)	0.18 (0.02)	3.4 (1.8)	−1.1 (1.1)
$RFC_{OCC\text{-}PM,spa}$	Maximum depth = 5, maximum features = square root	0.67 (0.10)	0.55 (0.07)	0.18 (0.03)	3.7 (2.6)	−1.4 (1.8)
$RFC_{MSC\text{-}PM,spa}$	Maximum depth = 5, maximum features = square root	0.67 (0.08)	0.56 (0.07)	0.19 (0.03)	3.4 (2.1)	−1.2 (1.4)

PLR, penalised logistic regression; RFC, random forest classification; l1, LASSO regularisation; l2, ridge regularisation; OCC-PM, combined oral cavity and pharyngeal mucosa (extending inferiorly to level of inferior border of oral cavity) contours; MSC-PM, combined oral mucosal surfaces and pharyngeal mucosa (extending inferiorly to level of inferior border of oral mucosal surfaces) contours; sta, standard model (dose-volume metrics only); spa, spatial model (dose-volume plus spatial dose metrics).

Table 2
Regression coefficients and covariate transformation values for $PLR_{OCC\text{-}PM,sta}$ and $PLR_{MSC\text{-}PM,sta}$ models

Covariate	Regression coefficient		Mean	Standard deviation
	$PLR_{OCC\text{-}PM,sta}$	$PLR_{MSC\text{-}PM,sta}$		
Intercept	0.000	0.002	–	–
definitiveRT	0.000	−0.009	0.832	0.374
Male	0.000	0.007	0.637	0.481
Age	0.000	−0.006	56.7	11.9
indChemo	0.000	0.010	0.497	0.500
noConChemo	0.000	−0.012	0.553	0.497
Cisplatin	0.000	0.013	0.358	0.479
Carboplatin	0.000	0.002	0.0559	0.230
cisCarbo	0.000	−0.004	0.0335	0.180
hypopharynx/larynx	0.000	0.005	0.101	0.301
Oropharynx	0.000	0.021	0.559	0.497
Nasopharynx	0.000	0.005	0.101	0.301
unknown primary	−0.012	−0.011	0.0447	0.207
Parotid	−0.243	−0.028	0.196	0.397
V020	0.000	−0.006	96.2/91.8	9.72/16.1
V040	0.000	−0.004	94.6/88.3	12.2/18.3
V060	0.000	0.004	92.6/83.1	14.1/20.0
V080	0.000	0.011	88.2/76.6	17.1/21.7
V100	0.000	0.017	82.5/70.1	21.3/22.4
V120	0.000	0.020	76.8/63.4	24.4/22.3
V140	0.000	0.021	70.9/56.8	25.9/21.6
V160	0.000	0.023	64.5/50.0	26.5/20.4
V180	0.201	0.022	57.2/43.2	26.4/19.4
V200	0.000	0.023	47.4/35.2	27.1/19.6
V220	0.250	0.025	12.4/11.1	11.4/9.41
V240	0.000	0.003	0.170/0.303	1.24/2.16
V260	0.000	0.000	0.000/0.000	1.00/1.00

The regression coefficients of many of the covariates in the $PLR_{OCC\text{-}PM,sta}$ model were set to 0 by the LASSO penalisation. Vx, volume of organ receiving x cGy of radiation per fraction. Mean and standard deviation for dose-volume metrics, Vx are given as OCC-PM/MSC-PM. definitiveRT, definitive radiation therapy (versus postoperative radiation therapy); indChemo, induction chemotherapy; noConChemo, no concurrent chemotherapy; cisCarbo, one cycle of cisplatin followed by one cycle of carboplatin.

Please cite this article in press as: Dean JA, et al., Normal Tissue Complication Probability (NTCP) Modelling of Severe Acute Mucositis using a Novel Oral Mucosal Surface Organ at Risk, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.12.001

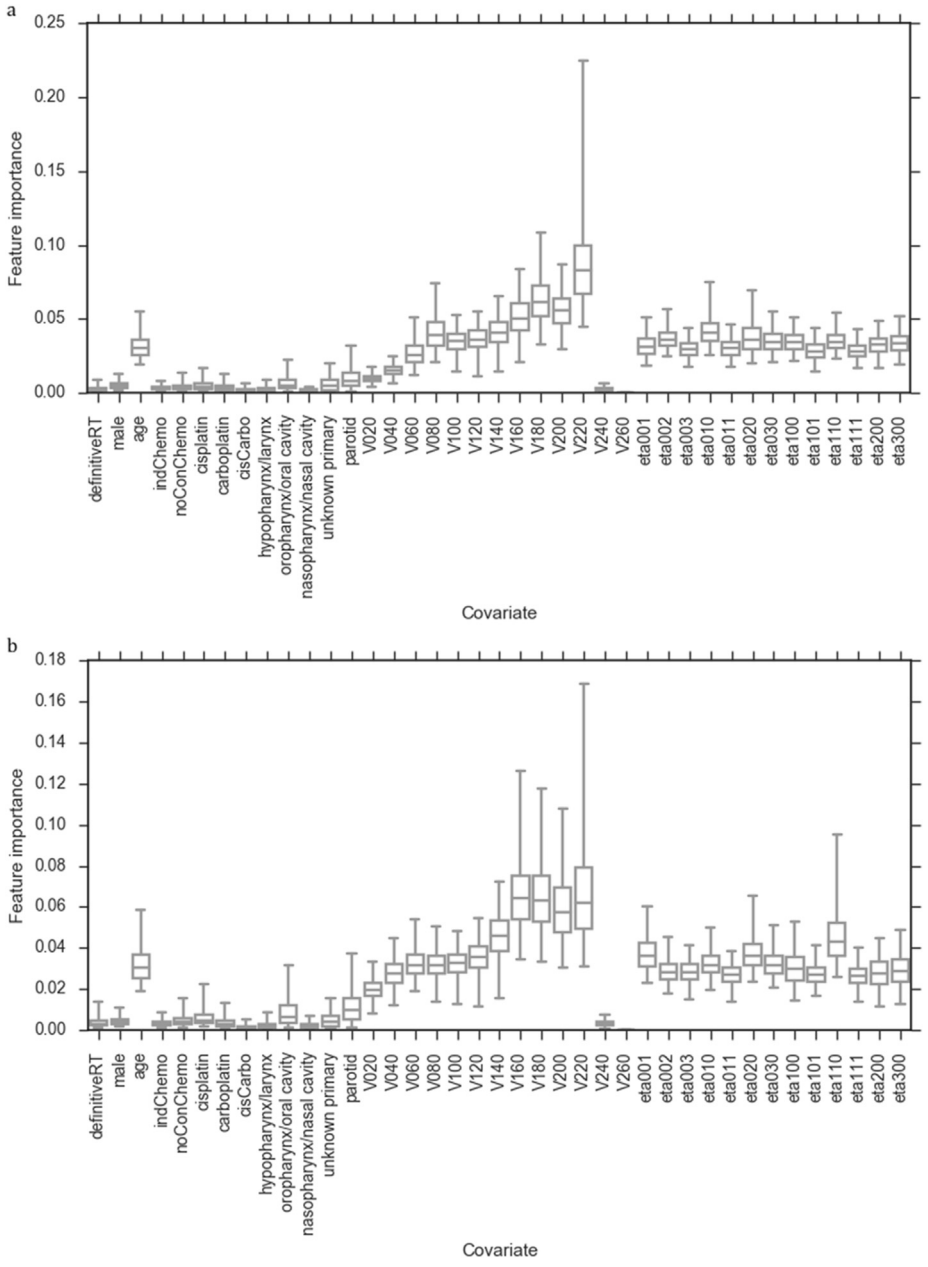

Fig 3. Random forest classification feature importance values for the (a) $RFC_{OCC\text{-}PM,spa}$ and (b) $RFC_{MSC\text{-}PM,spa}$ models, bootstrapped with 2000 replicates. The whiskers show the 95 percentile confidence intervals.

these covariates in our dataset (as the values of these dose metrics approach zero for all patients). The spatial dose metric with the highest feature importance was η_{110}, which represents the centre of mass of the dose in the diagonal left–right–anterior–posterior direction. Age was the clinical covariate with the highest feature importance in both models. However, it should be noted that this may be related to the fact that RFC feature importance can be artificially inflated for covariates having a large number of different values [42]. Age was negatively correlated with severe mucositis on univariable (Appendix E) and multivariable (Table 2; small negative regression coefficient) analysis.

Discussion

We generated models of severe acute mucositis using a novel MSC OAR and spatial dose metrics with the aims of: (i) improving the ability of NTCP models to predict which patients would experience severe mucositis and (ii) improving our understanding of the link between the

Please cite this article in press as: Dean JA, et al., Normal Tissue Complication Probability (NTCP) Modelling of Severe Acute Mucositis using a Novel Oral Mucosal Surface Organ at Risk, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.12.001

radiotherapy dose distribution and severe mucositis. We determined that using our novel MSC OAR to describe the dose distribution did not improve predictive performance over our previously used OCC OAR. This is probably due to the fact that, although the dose metrics between the two contouring approaches were different, they were highly correlated (Appendix E). We therefore recommend the use of the simpler OCC OAR for developing mucositis severity (not including spatial extent) clinical decision-support tools. The RFC models should be favoured over the PLR models for their ability to make superior probability predictions and only slightly inferior discrimination. Moreover, we established that adding spatial dose metrics, describing the centre of mass, spread and skewness of the dose distribution in three dimensions, did not improve the predictive performance of the models, compared with models using only dose-volume-based descriptors of the radiotherapy dose distribution. This is probably due to the fact that the CTCAE mucositis scoring instrument used in the clinical trials does not capture information on the location or extent of mucositis. We suggest that investigators consider the use of a mucositis scoring instrument sensitive to the spatial extent of mucositis, such as the Oral Mucositis Assessment Scale [43], for the study of radiation-induced mucositis in the future. We expect that the MSC OAR and spatial dose metrics may improve the discriminative ability of models of mucositis extent. However, this requires evaluation.

With regards to the second aim, we determined that the volumes of oral mucosa receiving high and intermediate doses (about 160 cGy per fraction and higher) were the treatment covariates with the strongest associations with mucositis severity. Although this is intuitive, mean oral cavity dose is currently used as an objective in radiotherapy planning protocols (for example in the RTOG 0912, RTOG 0920 and RTOG 1216 trials). The mean dose gives equal importance weighting to low doses as high doses. We recommend that protocols attempt to reduce the volumes of oral mucosa receiving high and intermediate dose levels instead.

Our group has previously published models of severe acute mucositis [13,44], which we attempted to improve upon in this study. We recommend our previous model [13] be favoured over the models in this study, as it was not outperformed by any of those described here. To the best of our knowledge, no other (internally or externally) validated models of severe acute mucositis, allowing individualised risk estimates to be made (aim i), have been published in the intensity-modulated radiotherapy era. Conversely, several studies have focussed on establishing the covariates that were most strongly associated with toxicity (aim ii). In support of our findings of associations between radiotherapy dose and mucositis, a small prospective study showed that cumulative point doses of 32 Gy or lower (with varying fractionations) were associated with minimal mucositis at those points [27] and a larger study found associations between weekly doses of 10.1 Gy and severe mucositis [45]. Also, a small randomised trial has shown the benefits of sparing the oral mucosa outside of the planning target volume (in patients with oral tongue cancers), when the spatial location of mucositis was included in the toxicity scoring system [46]. Although there has been some evidence of increased mucositis with the use of concurrent chemotherapy [45,47], our results, in agreement with several other studies [48,49], did not support a strong association between concurrent chemotherapy and severe mucositis.

Our study possesses several limitations. We did not have access to data pertaining to all of the covariates (for example, genetic polymorphisms [50], tobacco and alcohol) in the published studies and, so, were unable to attempt to replicate all of their findings using our data. Moreover, competing dynamic biological processes govern the onset, progression and resolution of mucositis [28]. Substantial improvements in predictive performance may require relevant data relating to those processes. We recommend that, in the future, investigators should endeavour to generate NTCP models featuring both relevant biological data (for example [51]) and a sufficiently detailed description of the radiotherapy dose distribution (such as that used in this study). Furthermore, the dose distribution was described using the dose calculated on planning computed tomography scans, which is an approximation of the delivered dose, probably featuring inaccuracies due to interfraction motion and patient positioning errors. Image-guided radiotherapy images were unavailable for many of the patients in this study so dose accumulation [52] could not be carried out.

Conclusions

We determined that using a novel MSC OAR did not improve the predictive performance of severe acute mucositis NTCP models. Exploring dose-response associations using both OCC and MSC OARs led us to recommend that radiotherapy planning protocols should prioritise the reduction of the volumes of oral mucosa receiving high and intermediate doses, rather than reducing the mean dose.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council, Cancer Research UK Programme Grant A13407 and NHS funding to the NIHR Biomedical Research Centre at The Royal Marsden and ICR. The PARSPORT and COSTAR trials were supported by Cancer Research UK (trial reference numbers CRUK/03/005 and CRUK/08/004). None of the funding sources had any role in any part of the study or manuscript preparation and submission. We wish to thank Hannah Eyles, Emma Wells, James Morden and Dr Emma Hall at The Institute of Cancer Research Clinical Trials and Statistics Unit for data collation, Dr Cornelis Kamerling, Dr Alex Dunlop, Dr Dualta McQuaid, Dr Simeon Nill and Professor Uwe Oelfke for general support and Dr Jung Hun Oh and Professor Joseph Deasy for insightful discussions.

Please cite this article in press as: Dean JA, et al., Normal Tissue Complication Probability (NTCP) Modelling of Severe Acute Mucositis using a Novel Oral Mucosal Surface Organ at Risk, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.12.001

Appendix A. Patient Data

Table A1
Clinical trials making up the dataset

Trial	Patients available	Primary disease site	Radiation therapy technique	Radiation therapy dose-fractionation	Concurrent chemotherapy
COSTAR (phase III, multicentre) [23]	78	Parotid gland	Unilateral; conventional, IMRT	65 Gy/30 fractions (definitive radiotherapy), 60 Gy/30 fractions (postoperative radiotherapy)	No
PARSPORT (phase III, multicentre) [17]	71	Oropharynx, hypopharynx	Bilateral; conventional IMRT	65 Gy/30 fractions (definitive radiotherapy), 60 Gy/30 fractions (postoperative radiotherapy)	No
Dose escalation (phase II, single centre) [18,21]	30	Larynx, hypopharynx	Bilateral; IMRT	67.2 Gy/28 fractions, 63 Gy/28 fractions	Yes
Midline (phase II, single centre) [19]	117	Oropharynx	Bilateral; IMRT	65 Gy/30 fractions (definitive radiotherapy), 60 Gy/30 fractions (postoperative radiotherapy)	Yes
Nasopharynx (phase II, single centre) [20]	36	Nasopharynx	Bilateral; IMRT	65 Gy/30 fractions (definitive radiotherapy), 60 Gy/30 fractions (postoperative radiotherapy)	Yes
Unknown primary (phase II, single centre) [22]	19	Unknown primary	Bilateral; IMRT	65 Gy/30 fractions (definitive radiotherapy), 60 Gy/30 fractions (postoperative radiotherapy)	Yes

IMRT, intensity-modulated radiotherapy; unilateral, treatment delivered to ipsilateral parotid bed only; bilateral, treatment delivered to ipsilateral and contralateral mucosa of relevant subsite (e.g. nasopharynx, oropharynx or larynx).

Appendix B. Strategy for Handling Missing Data

If weekly toxicity data are incomplete this can lead to assignment of an incorrect peak toxicity grade. For example, consider a patient who has grade 1 toxicity for weeks 1–3, grade 2 toxicity for weeks 4 and 5, missing toxicity week 6 and 1 week after treatment and grade 2 toxicity from 2 weeks after radiotherapy to 8 weeks after radiotherapy. They would be assigned a peak grade of 2. However, they may, in fact, have experienced grade 3 toxicity, which was not scored, as they were unable to attend their follow-up appointments. This would introduce an error into the analysis. As this type of error can only lead to peak toxicity being under-scored and not over-scored it could introduce bias. Therefore, in an attempt to reduce bias at the expense of statistical power, patients with any missing toxicity scores and a peak score below 3 were excluded from the analysis. Missing toxicity data were not imputed as many patients (with full toxicity data) with peak toxicity of grade 3 were only scored as grade 3 for 1 week. We previously investigated the effects of imputing missing toxicity measurements, where there were non-consecutive missing values and found that this made little difference [13]. Patients with some missing toxicity measurements, but at least one measurement scored as grade 3 were included as they must have a peak grade of 3 or higher. It should be noted that retaining patients with missing data, but having a peak grade of 3 skews the apparent incidences of peak toxicity grades. There was a general trend that data were more likely to be missing around the middle to the end of treatment, which was when the peak grade of toxicity tended to occur.

It should be noted that our approach to handling missing data might still result in bias. Where there are missing data there is always a risk of bias, whichever method for handling missing data is used. This is particularly true where the data are not missing at random, as is suggested by the pattern of missing data in this dataset. Ultimately, the performance of the model, including any bias introduced by the missing data handling strategy, should be assessed by external validation.

Appendix C. Example of Three-dimensional Moment Invariants

Figure C1 shows oral cavity dose distributions for one patient receiving bilateral irradiation and another receiving unilateral irradiation.

Please cite this article in press as: Dean JA, et al., Normal Tissue Complication Probability (NTCP) Modelling of Severe Acute Mucositis using a Novel Oral Mucosal Surface Organ at Risk, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.12.001

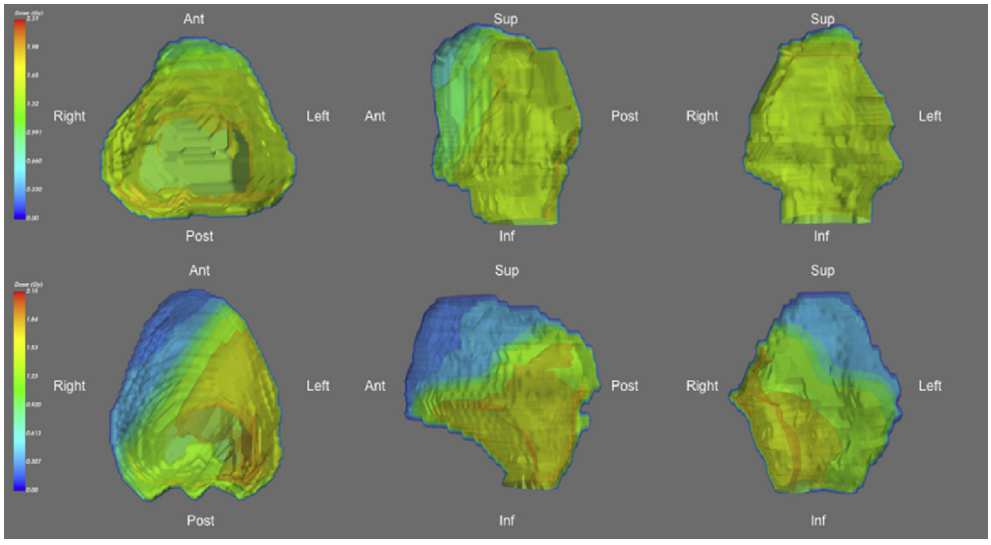

Fig C1. Example oral cavity dose distributions for patients receiving bilateral (top) and unilateral (bottom) irradiation. Sup, superior; inf, inferior; ant, anterior; post, posterior. The corresponding three-dimensional moment invariants for these patients are given in Figure C2.

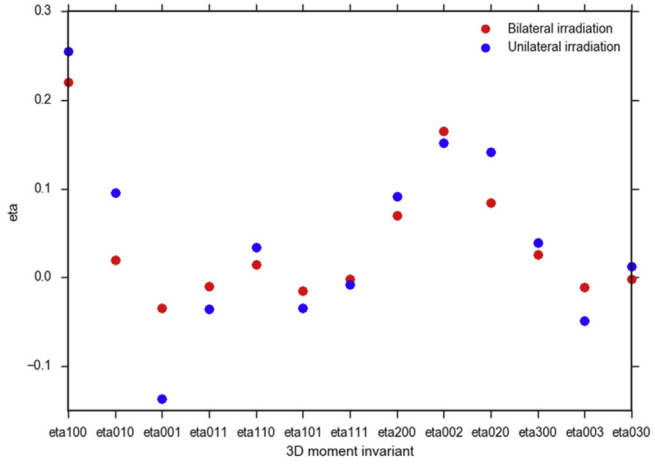

Fig C2. Example three-dimensional moment invariants for patients receiving bilateral (red) and unilateral (blue) irradiation.

Appendix D. Hyper-parameters for Cross-validated Grid Search

- PLR: regularisation = {LASSO (l1), ridge (l2)}; inverse regularisation strength (C) = {0.001, 0.01, 0.1, 1.0, 10.0, 100.0, 1000.0}.
- RFC: number of estimators = 1000; maximum depth = {5, 10, 15, 20}; maximum features = {number of features, number of features/2, square root of number of features}.

Appendix E. Correlation Matrix

Figure E1 shows the correlation matrix of the covariates and mucositis severity.

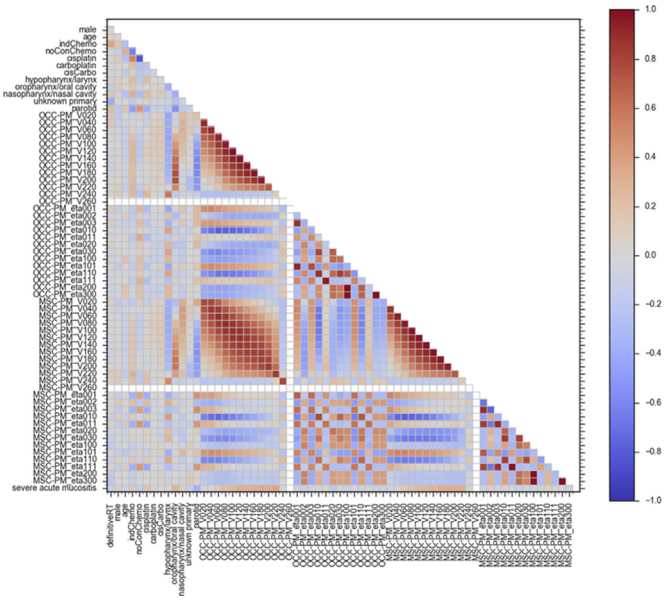

Fig E1. Correlation matrix of the variables included in the modelling. The colour scale shows the Spearman correlation coefficients between the variables. definitiveRT, definitive radiation therapy (versus postoperative radiation therapy); indChemo, induction chemotherapy; noConChemo, no concurrent chemotherapy; cisCarbo, one cycle of cisplatin followed by one cycle of carboplatin; OCC-PM, oral cavity contours and pharyngeal mucosa organ at risk; MSC-PM, mucosal surface contours and pharyngeal mucosa organ at risk; Vx, volume of organ receiving x cGy of radiation per fraction; etaxyz, three-dimensional moment invariant

Please cite this article in press as: Dean JA, et al., Normal Tissue Complication Probability (NTCP) Modelling of Severe Acute Mucositis using a Novel Oral Mucosal Surface Organ at Risk, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.12.001

of order x in the left–right direction, y in the anterior–posterior direction and z in superior–inferior direction; severe acute mucositis, peak acute mucositis severity (non-severe = 0, severe = 1).

References

[1] Sanguineti G, Gunn GB, Parker BC, Endres EJ, Zeng J, Fiorino C. Weekly dose-volume parameters of mucosa and constrictor muscles predict the use of percutaneous endoscopic gastrostomy during exclusive intensity-modulated radiotherapy for oropharyngeal cancer. *Int J Radiat Oncol Biol Phys* 2011;79:52–59. http://dx.doi.org/10.1016/j.ijrobp.2009.10.057.

[2] Trotti A. Toxicity in head and neck cancer: a review of trends and issues. *Int J Radiat Oncol Biol Phys* 2000;47:1–12. http://dx.doi.org/10.1016/S0360-3016(99)00558-1.

[3] Kelly C, Paleri V, Downs C, Shah R. Deterioration in quality of life and depressive symptoms during radiation therapy for head and neck cancer. *Otolaryngol Head Neck Surg* 2007;136:108–111. http://dx.doi.org/10.1016/j.otohns.2006.06.1278.

[4] Trotti A, Bellm LA, Epstein JB, *et al*. Mucositis incidence, severity and associated outcomes in patients with head and neck cancer receiving radiotherapy with or without chemotherapy: a systematic literature review. *Radiother Oncol* 2003;66:253–262. http://dx.doi.org/10.1016/S0167-8140(02)00404-8.

[5] Bourhis J, Overgaard J, Audry H, *et al*. Hyperfractionated or accelerated radiotherapy in head and neck cancer: a meta-analysis. *Lancet* 2006;368:843–854. http://dx.doi.org/10.1016/S0140-6736(06)69121-6.

[6] Jackson SM, Weir LM, Hay JH, Tsang VH, Durham JS. A randomised trial of accelerated versus conventional radiotherapy in head and neck cancer. *Radiother Oncol* 1997;43:39–46. http://dx.doi.org/10.1016/S0167-8140(97)01944-0.

[7] Maciejewski B, Skladowski K, Pilecki B, *et al*. Randomized clinical trial on accelerated 7 days per week fractionation in radiotherapy for head and neck cancer. Preliminary report on acute toxicity. *Radiother Oncol* 1996;40:137–145. http://dx.doi.org/10.1016/0167-8140(96)01776-8.

[8] Sonis ST. Mucositis: the impact, biology and therapeutic opportunities of oral mucositis. *Oral Oncol* 2009;45:1015–1020. http://dx.doi.org/10.1016/j.oraloncology.2009.08.006.

[9] Bentzen SM. Preventing or reducing late side effects of radiation therapy: radiobiology meets molecular pathology. *Nat Rev Cancer* 2006;6:702–713. http://dx.doi.org/10.1038/nrc1950.

[10] Denham JW, Peters LJ, Johansen J, *et al*. Do acute mucosal reactions lead to consequential late reactions in patients with head and neck cancer? *Radiother Oncol* 1999;52:157–164. http://dx.doi.org/10.1016/S0167-8140(99)00107-3.

[11] Lambin P, van Stiphout RGPM, Starmans MHW, *et al*. Predicting outcomes in radiation oncology – multifactorial decision support systems. *Nat Rev Clin Oncol* 2013;10:27–40. http://dx.doi.org/10.1038/nrclinonc.2012.196.

[12] Sanguineti G, Endres EJ, Gunn BG, Parker B. Is there a "mucosa-sparing" benefit of IMRT for head-and-neck cancer? *Int J Radiat Oncol Biol Phys* 2006;66:931–938. http://dx.doi.org/10.1016/j.ijrobp.2006.05.060.

[13] Dean JA, Wong KH, Welsh LC, *et al*. Normal tissue complication probability (NTCP) modelling using spatial dose metrics and machine learning methods for severe acute oral mucositis resulting from head and neck radiotherapy. *Radiother Oncol* 2016;120:21–27. http://dx.doi.org/10.1016/j.radonc.2016.05.015.

[14] Bhide SA, Gulliford S, Fowler J, *et al*. Characteristics of response of oral and pharyngeal mucosa in patients receiving chemo-IMRT for head and neck cancer using hypofractionated accelerated radiotherapy. *Radiother Oncol* 2010;97:86–91. http://dx.doi.org/10.1016/j.radonc.2010.08.013.

[15] Dean JA, Welsh LC, Gulliford SL, Harrington KJ, Nutting CM. A novel method for delineation of oral mucosa for radiotherapy dose-response studies. *Radiother Oncol* 2015;115:63–66. http://dx.doi.org/10.1016/j.radonc.2015.02.020.

[16] Dean JA, Welsh LC, McQuaid D, *et al*. Assessment of fully-automated atlas-based segmentation of novel oral mucosal surface organ-at-risk. *Radiother Oncol* 2016;119:166–171. http://dx.doi.org/10.1016/j.radonc.2016.02.022.

[17] 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. http://dx.doi.org/10.1016/S1470-2045(10)70290-4.

[18] Miah AB, Bhide SA, Guerrero-Urbano MT, *et al*. Dose-escalated intensity-modulated radiotherapy is feasible and may improve locoregional control and laryngeal preservation in laryngo-hypopharyngeal cancers. *Int J Radiat Oncol Biol Phys* 2012;82:539–547. http://dx.doi.org/10.1016/j.ijrobp.2010.09.055.

[19] Miah AB, Schick U, Bhide SA, Clark CH, Bidmead AM, Bodla S. A phase II trial of induction chemotherapy and chemo-IMRT for head and neck squamous cell cancers at risk of bilateral nodal spread: the application of a bilateral superficial lobe parotid-sparing IMRT technique and treatment outcomes. *Br J Cancer* 2014:1–7. http://dx.doi.org/10.1038/bjc.2014.553.

[20] Miah AB, Bhide SA, Del Rosario L, *et al*. Induction chemotherapy followed by chemo-intensity-modulated radiotherapy for locally advanced nasopharyngeal cancer. *Clin Oncol* 2016;28:e1–e7. http://dx.doi.org/10.1016/j.clon.2016.01.012.

[21] Gujral DM, Miah AB, Bodla S, *et al*. Final long-term results of a phase I/II study of dose-escalated intensity-modulated radiotherapy for locally advanced laryngo-hypopharyngeal cancers. *Oral Oncol* 2014;50:1089–1097. http://dx.doi.org/10.1016/j.oraloncology.2014.07.018.

[22] Richards TM, Bhide SA, Miah AB, *et al*. Total mucosal irradiation with intensity-modulated radiotherapy in patients with head and neck carcinoma of unknown primary: a pooled analysis of two prospective studies. *Clin Oncol* 2016;28:e77–e84. http://dx.doi.org/10.1016/j.clon.2016.04.035.

[23] Nutting C, Morden JP, Beasley M, *et al*. First results of COSTAR: a randomised trial of 3-dimensional conformal radiotherapy (3DCRT) vs cochlea-sparing intensity modulated radiotherapy (CS-IMRT) in patients with parotid cancer. *J Clin Oncol* 2016;34 (suppl):6006.

[24] The National Cancer Institute. *Common Toxicity Criteria (CTC) Version 2.0* 1999.

[25] The National Cancer Institute. *Common Terminology Criteria for Adverse Events v3.0 (CTCAE)* 2006.

[26] Tucker SL, Michalski JM, Bosch WR, *et al*. Use of fractional dose-volume histograms to model risk of acute rectal toxicity among patients treated on RTOG 94-06. *Radiother Oncol* 2012;104:109–113. http://dx.doi.org/10.1016/j.radonc.2012.04.023.

[27] Narayan S, Lehmann J, Coleman MA, *et al*. Prospective evaluation to establish a dose response for clinical oral mucositis in patients undergoing head-and-neck conformal radiotherapy. *Int J Radiat Oncol Biol Phys* 2008;72:756–762. http://dx.doi.org/10.1016/j.ijrobp.2008.01.060.

[28] Sonis ST. The pathobiology of mucositis. *Nat Rev Cancer* 2004;4:277–284. http://dx.doi.org/10.1038/nrc1318.

[29] Buettner F, Miah AB, Gulliford SL, *et al*. Novel approaches to improve the therapeutic index of head and neck radiotherapy: an analysis of data from the PARSPORT randomised

Please cite this article in press as: Dean JA, et al., Normal Tissue Complication Probability (NTCP) Modelling of Severe Acute Mucositis using a Novel Oral Mucosal Surface Organ at Risk, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.12.001

phase III trial. *Radiother Oncol* 2012;103:82–87. http://dx.doi.org/10.1016/j.radonc.2012.02.006.
[30] Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 1970;12:55–67. http://dx.doi.org/10.1080/00401706.1970.10488634.
[31] Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B* 1996;58:267–288. http://dx.doi.org/10.1111/j.1467-9868.2011.00771.x.
[32] Breiman L. Random forests. *Mach Learn* 2001;45:5–32. http://dx.doi.org/10.1023/A:1010933404324.
[33] Moons KGM, Altman DG, Reitsma JB, *et al*. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1–W73. http://dx.doi.org/10.7326/M14-0698.
[34] Moons KGM, de Groot JAH, Bouwmeester W, *et al*. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014;11. http://dx.doi.org/10.1371/journal.pmed.1001744.
[35] Brier GW. Verification of forecasts expressed in terms of probaility. *Mon Weather Rev* 1950;78:1–3. http://dx.doi.org/10.1126/science.27.693.594.
[36] Good IJ. Rational decisions. *J R Stat Soc Ser B* 1952;14:107–114.
[37] Steyerberg EW, Vickers AJ, Cook NR, *et al*. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–138. http://dx.doi.org/10.1097/EDE.0b013e3181c30fb2.
[38] Pavlou M, Ambler G, Seaman SR, *et al*. How to develop a more accurate risk prediction model when there are few events. *Br Med J* 2015;351:h3868. http://dx.doi.org/10.1136/bmj.h3868.
[39] Rossum G. *Python Reference Manual* 1995.
[40] McKinney W. Data structures for statistical computing in Python. *Proc 9th Python Sci Conf* 2010:51–56.
[41] Pedregosa F, Varoquaux G, Gramfort A, *et al*. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;12:2825–2830.
[42] Strobl C, Boulesteix A-L, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics* 2007;8:25. http://dx.doi.org/10.1186/1471-2105-8-25.
[43] Sonis ST, Eilers JP, Epstein JB, *et al*. Validation of a new scoring system for the assessment of clinical trial research of oral mucositis induced by radiation or chemotherapy. *Cancer* 1999;85:2103–2113. http://dx.doi.org/10.1002/(SICI)1097-0142(19990515)85:10<2103::AID-CNCR2>3.0.CO;2–0.
[44] Otter S, Schick U, Gulliford S, *et al*. Evaluation of the risk of grade 3 oral and pharyngeal dysphagia using atlas-based method and multivariate analyses of individual patient dose distributions. *Int J Radiat Oncol Biol Phys* 2015;93:507–515. http://dx.doi.org/10.1016/j.ijrobp.2015.07.2263.
[45] Sanguineti G, Sormani MP, Marur S, *et al*. Effect of radiotherapy and chemotherapy on the risk of mucositis during intensity-modulated radiation therapy for oropharyngeal cancer. *Int J Radiat Oncol Biol Phys* 2012;83:235–242. http://dx.doi.org/10.1016/j.ijrobp.2011.06.2000.
[46] Wang Z-H, Zhang S-Z, Zhang Z-Y, *et al*. Protecting the oral mucosa in patients with oral tongue squamous cell carcinoma treated postoperatively with intensity-modulated radiotherapy: a randomized study. *Laryngoscope* 2012;122:291–298. http://dx.doi.org/10.1002/lary.22434.
[47] Vera-Llonch M, Oster G, Hagiwara M, Sonis S. Oral mucositis in patients undergoing radiation treatment for head and neck carcinoma. *Cancer* 2006;106:329–336. http://dx.doi.org/10.1002/cncr.21622.
[48] Elting LS, Keefe DM, Sonis ST, *et al*. Patient-reported measurements of oral mucositis in head and neck cancer patients treated with radiotherapy with or without chemotherapy: demonstration of increased frequency, severity, resistance to palliation, and impact on quality of life. *Cancer* 2008;113:2704–2713. http://dx.doi.org/10.1002/cncr.23898.
[49] Epstein JB, Beaumont JL, Gwede CK, *et al*. Longitudinal evaluation of the oral mucositis weekly questionnaire-head and neck cancer, a patient-reported outcomes questionnaire. *Cancer* 2007;109:1914–1922. http://dx.doi.org/10.1002/cncr.22620.
[50] Werbrouck J, De Ruyck K, Duprez F, *et al*. Acute normal tissue reactions in head-and-neck cancer patients treated with IMRT: influence of dose and association with genetic polymorphisms in DNA DSB repair genes. *Int J Radiat Oncol Biol Phys* 2009;73:1187–1195. http://dx.doi.org/10.1016/j.ijrobp.2008.08.073.
[51] Sonis S, Antin J, Tedaldi M, Alterovitz G. SNP-based Bayesian networks can predict oral mucositis risk in autologous stem cell transplant recipients. *Oral Dis* 2013;19:721–772. http://dx.doi.org/10.1111/odi.12146.
[52] Jaffray DA, Lindsay PE, Brock KK, Deasy JO, Tome WA. Accurate accumulation of dose for improved understanding of radiation effects in normal tissue. *Int J Radiat Oncol Biol Phys* 2010;76:135–139. http://dx.doi.org/10.1016/j.ijrobp.2009.06.093.

Please cite this article in press as: Dean JA, et al., Normal Tissue Complication Probability (NTCP) Modelling of Severe Acute Mucositis using a Novel Oral Mucosal Surface Organ at Risk, Clinical Oncology (2016), http://dx.doi.org/10.1016/j.clon.2016.12.001