Title

Heterogeneous radiotherapy dose-outcomes response in parotid glands

Journal

Convergent Science Physical Oncology, 4(3)

ISSN

2057-1739

Authors

Clark, Haley D
Thomas, Steven D
Reinsberg, Stefan A
et al.

Publication Date

2018-06-18

DOI

10.1088/2057-1739/aac8ea

Peer reviewed

Convergent Science Physical Oncology

PAPER

RECEIVED
1 February 2018

REVISED
7 May 2018

ACCEPTED FOR PUBLICATION
31 May 2018

PUBLISHED
18 June 2018

Heterogeneous radiotherapy dose-outcomes response in parotid glands

Haley D Clark[1,7], Steven D Thomas[1], Stefan A Reinsberg[2], Vitali V Moiseenko[3], Allan J Hovan[4] and Jonn S Wu[5,6]

[1] Department of Medical Physics, British Columbia Cancer Agency, BC, Canada
[2] Department of Physics and Astronomy, University of British Columbia, BC, Canada
[3] Department of Radiation Medicine and Applied Sciences, University of California San Diego, CA, United States of America
[4] Department of Oral Oncology, British Columbia Cancer Agency, BC, Canada
[5] Department of Radiation Oncology, British Columbia Cancer Agency, BC, Canada
[6] Department of Surgery, University of British Columbia, BC, Canada

E-mail: halclark.ca/Contact.html

Keywords: heterogeneous functional burden, regional effects, salivary dysfunction, parotid gland, relative importance

Supplementary material for this article is available online

Abstract

Parotid glands are treated clinically as though the distribution of functional burden were homogeneous. Radiotherapy treatments are planned using whole parotid mean dose to predict risk of salivary dysfunction. Recent progress has identified specific parotid non-homogeneities by demonstrating the existence of regional, bath-and-shower, and dose-volume effects. In this work, parotid regional effects and their impact on salivary function are quantified using a non-parametric (model-free) approach. Regional effects have implications for clinical sparing practices. Radiotherapy planning contours, dose profiles, and late clinical outcomes from a single cohort consisting of $N = 332$ patients was used. Pre-radiotherapy and one year post-radiotherapy whole mouth stimulated saliva were collected for assessment of salivary dysfunction. Organ-at-risk parotid glands were segmented into 2, 3, 4, 18, and 96 equal-volume sub-segments. Sub-segment relative importance was derived from mean dose regressors using random forests and conditional inference trees. Regressor multicollinearity, cohort homogeneity, and overfitting were addressed. Linear and exponential whole parotid mean dose models were also implemented for comparison purposes. Exclusion of caudal-anterior sub-segments negatively impacted prediction the most. The most important sub-segments had importances 2.4× (on average over all segmentation methods) or >4× (at the finest level of segmentation) that of an equivalent sub-segment in a theoretical homogeneous parotid. In contrast, the least important sub-segments held virtually no importance for prediction. Both random forests and conditional inference trees outperformed parametric (model-based) techniques. Both improved prediction as segmentation was refined. Radiation dose to caudal-anterior aspects of the parotid are the strongest predictors of radiotherapy-induced late stimulated whole mouth saliva, and are thus the most clinically-relevant regions for controlling dysfunction. Cranial and posterior aspects are less important. Shifting dose from regions of high importance to low importance may therefore improve patient outcomes.

1. Introduction

Whole parotid mean radiation dose is currently used to predict risk of late radiotherapy-induced salivary dysfunction [1]. The underlying assumption is that functional burden is distributed homogeneously throughout the parotid gland [2]. Recent studies have found behaviour counter to homogeneous distribution, including regions with elevated relevance for salivary flow [3, 4], non-equivalence of dose-volume descriptors for dysfunction prediction [5, 6], and bath-and-shower effects [7, 8]. Others have noted

[7] Author to whom any correspondence should be addressed.

that incorporation of non-homogeneous effects into a radiotherapy treatment plan leads or potentially could lead to improved patient outcomes [9, 10].

Evidence of a bath-and-shower effect in parotid, wherein the effect of high dose to a confined sub-volume ('shower') is impacted by a low dose to an extended volume ('bath'), was first reported by van Luijk *et al* in 2009 in the context of objective salivary flow dysfunction [7]. Specifically, the addition of a bath dose $<$10Gy to a shower dose to the caudal half resulted in a disproportionally high enhancement of dysfunction. van Luijk *et al* conjectured the bath-and-shower effect might explain the outcomes shortfall experienced when switching from conventional radiotherapy to modulated therapies. A similar effect was found using a separate cohort and subjective measurements in 2012 [8]. Likewise, several studies have confirmed that dose-volume measures are not equivalent in parotid, implying deviation from homogeneity. For example, Ortholan *et al* found that salivary flow prediction improved compared to whole mean dose models when the volume of the contralateral gland receiving $\geqslant$40 Gy was incorporated [5]. Wang *et al* found similar conclusions in 2011 [6]. However, neither dose-volume effect deviations nor bath-and-shower effects incorporate *specific* sub-volumes; incorporation of sub-volume extent and location has lead to less conclusive findings. There is continued debate over the existence of critical regions (i.e. defined by specific anatomical, functional, or geographical criteria) that more strongly impact salivary dysfunction than comparable regions in the parotid. Different studies have variously shown that the most important regions are (or contain, or are contained broadly within) cranial and medial-dorsal aspects adjacent to mandible [3, 11], caudal-medial aspects [8], the superficial lobe (i.e. approximately lateral-caudal) [9], and the lateral-most half [4]. Other work has focused on the clinical feasibility of split delineation along the deep-superficial lobe boundary (i.e. anterolateral and posteromedial) [10, 12].

In this prospective study a cohort comprised of 332 head-and-neck cancer patients (collected within a single agency) is used to assess regional effects within parotid gland. Parotids are divided into equal-volume sub-segments and sub-segment relative importance for prediction of late salivary flow is assessed. Non-parametric methods robust to overfitting and multicollinearity are employed. Dose profiles in the cohort are ergodic in the sense that prescription doses, tumour site, and dose gradients across the parotid are heterogeneous. Owing to reports of linear dose-response (e.g. [13]) and sub-segment volumetric equality, importances are interpretable as regional criticality for late salivary dysfunction.

2. Materials and methods

2.1. Cohort, measurements, treatment, tooling

This prospective study passed institutional ethical review. Patients underwent radiotherapy for head-and-neck cancers and gave informed consent to participate. Planning dose profiles and delineated organ-at-risk parotid contours were employed for dosimetric assessment and segmentation. A single senior head and neck oncologist (JW) scrutinized contours for quality assurance. Stimulated salivary measurements of whole mouth saliva at baseline (pre-radiotherapy; W_b) and one year post-radiotherapy ('late'; W_{1y}) were used. Measurements represent whole mouth saliva. The saliva collection procedure described by Chao *et al* was employed [14]. In short, patients were asked to chew flavourless wax for a period of five minutes in a forward-leaning posture without swallowing. The saliva was directed into a small, pre-weighed cup and the mass of saliva was determined by weight. Mean-scaling imputation was employed for (29) patients without W_{1y} but with W_{2y} late measurements. Exclusion criteria are described in a supplementary document, available at stacks.iop.org/CSPO/4/035001/mmedia. A total of 332 patients were eligible (median age 58.6y, age range 19.0–90.6y; gender: 73% male, 27% female; prescription dose: 70Gy/35 fractions 55%, 60Gy/25 fractions 11%, 60Gy/35 fractions 8%, other 27%; treatment type: 279 intensity- or volumetric-modulated, 53 conventional; primary tumour site: 88 nasopharynx, 132 oropharynx, 61 tongue, 61 tonsil, 31 oral cavity and gums, 20 unknown, 18 hypopharynx, 14 larynx, 7 thyroid, 4 palate, and 22 other).

Parotids were divided into 2, 3, 4, 18, and 96 equal-volume sub-segments using nested segmentation. A depiction of nested segmentation applied to three parotids with differing morphologies is shown in figure 1. In brief, parotid contours are recursively partitioned along orthogonal planes. The plane directions are specified to align with sagittal, coronal, and transverse planes, but the positions are specified by the fraction of parotid volume cleaved by the plane. While the shape of sub-segments with an outer face adapts most distinctly to the outer parotid shape, the volume, position, and general shape of individual sub-segments is consistent regardless of gross morphology. Nested segmentation avoids the need to align parotids between patients or resort to a registration atlas. Contour manipulation in support of nested segmentation, and the assessment of radiation dose within each sub-segment was accomplished via `DICOMautomaton` [15, 16]. To ensure sub-segment dose was correctly estimated, cubic dose matrix supersampling (15$\times$) was employed. Counts of voxels within sub-segments were compared to ensure mutual pairwise proportionality using Kolmogorov–Smirnov tests. Significance was ascribed at α=0.05. No correction was made for multiple comparisons (i.e. to account for the so-called birthday paradox), which made for a *more stringent* test.

2.2. Importance techniques

The Random Forest technique (RF) is a non-parametric ensemble learning method in which

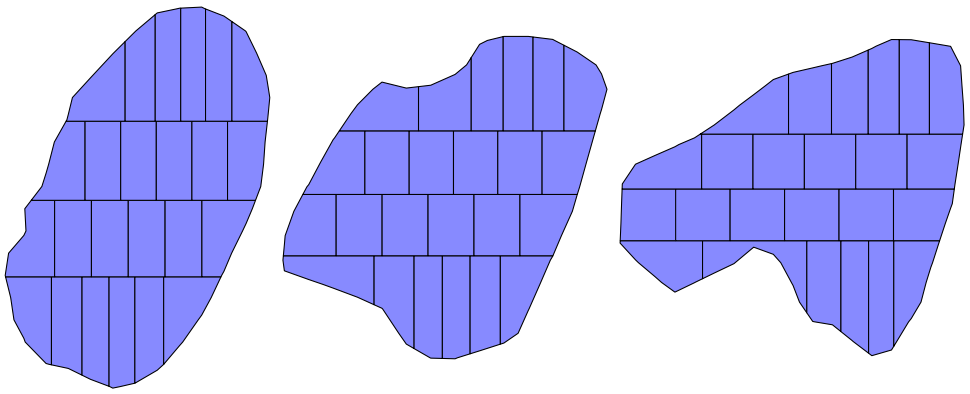

Figure 1. Nested segmentation applied to three distinctly-shaped parotid slices. This adaptive procedure results in iso-volumetric sub-segments with consistent position and shape within the parotid while also accounting for differing parotid morphology across the population. Exact adjacency is forfeited in order to reduce travel when the parotid contour is deformed. This results in robustness to contouring perturbations. The segmentation order was chosen to be straightforward to apply clinically.

tree nodes are recursively constructed by randomly sampling regressors at, and splitting, each node. An ensemble of trees is grown; regression predictions are generated by averaging predictions from the ensemble. Importance was estimated using two measures: (1) ensemble-averaged total decrease in node impurities resulting from splitting on the regressor and measuring the residual sum of squares (RSS) ('node impurity'), and (2) a more robust permutation-based measure in which the difference between un-permuted and each regressor permutation of the out-of-bag (i.e. excluded data) Mean squared error (MSE) is ensemble-averaged and normalized by the standard deviation of the differences (referred to as simply 'MSE' here) [17–19]. Major weaknesses of RF arise when regressors have varying scales, are mutually correlated ('multicollinearity'), or when the 'scale' (i.e. number of categories) of categorical variables differ [20]. In the present case all regressors (i.e. sub-segment mean doses) have the same scale and are continuous. Multicollinearity is anticipated, but is believed to be sufficiently pervasive and constant so as to reduce impact on conclusions by uniformly suppressing absolute importances and leaving relative importances intact.

RF trees may nonetheless become biased. To overcome this, conditional inference tree ensembles ('c-trees') were employed [21]. Like RF, c-trees can be used for non-parametric regression [22]. C-tree methods differ from RF by using conditional inference trees as base learners. The unbiased c-tree RF construction proposed by [20] is used, which is meant to address regressor selection bias in individual classification trees. Regressor importance is estimated using both (1) permutation and (2) conditional permutation measures. The former is a reliable measure of regressor importance for uncorrelated regressors when subsampling without replacement and unbiased trees are used to build the forest [20]. The latter, conditional permutation, is thought to be more suitable in the presence of multicollinearity and addresses regressor selection bias in individual classification trees [23].

Both RF and c-trees are thought to be robust to overfitting due to use of *bagging*, which is a bootstrap technique that improves generalizability [17, 22]. Based on expected multicollinearity, the reliability of importance estimates were ranked as: c-tree conditional permutation (most reliable), c-tree permutation, RF MSE, and RF node impurity (least reliable). The number of trees and splitting parameter were grown until impact on importances diminished and the random seed had no impact on conclusions (nominally 20000 for both RF and c-trees).

2.3. Statistics

Akaike's information criterion (AIC) is typically used to compare (parametric) models [24]. Besides an asymptotic relationship between cross-validation and AIC [25], the authors are not aware of any direct way to compute AIC for RF or c-trees. Instead, two metrics that characterize predictive power, mean absolute error (MAE) and root-mean-square error (RMSE), are reported [26, 27]. Fitted whole parotid mean dose models (linear and exponential; standard in the literature, e.g. [1] and [13]) provide baseline MAE and RMSE. Model fitting consists of approximating W_{1y}/W_b versus mean dose to the parotid with a straight line or exponential. The distribution of baseline-normalized salivary measurements will be heteroscedastic, so residual normality was not tested. Instead correlation coefficients (r_{pa}) between predicted and actual W_{1y}/W_b, are reported. Comparison is accomplished via a two-tailed Fischer z-transformation [28].

3. Results

A summary of all models and methods is shown in table 1. Contralateral parotid (i.e. the parotid with lowest mean dose) was unanimously more

Table 1. Summary of results and most importance sub-segments. All quantities are dimensionless. r_{pa} denotes the correlation coefficient between actual and predicted mean-scaled W_{1y}/W_b. Whole, halves, thirds, and quarters segmentation used both ipsi- and contralateral parotids; 18^{ths} and 96^{ths} used only contralateral parotids to reduce computational burden. The most important sub-segment (SS) is specified; refer to supplementary anatomical figures for locations. Importances given are relative to the expected result for a homogeneous parotid.

Segmentation	Method	MAE	RMSE	r_{pa}	Type	Most Important Sub-segment	Importance
Whole	exp	0.301	0.491	0.252	—	—	—
	linear	0.295	0.487	0.277	—	—	—
	RF	0.315	0.506	0.222	—	—	—
	c-trees	0.259	0.437	0.531	—	—	—
Halves	RF	0.294	0.488	0.272	Impurity	caudal (contralateral)	1.15×
					MSE	caudal (contralateral)	1.45×
	c-trees	0.246	0.425	0.591	Permutation	caudal (contralateral)	2.78×
					Conditional	caudal (contralateral)	2.66×
Thirds	RF	0.308	0.494	0.246	Impurity	caudal (contralateral)	1.31×
					MSE	caudal (contralateral)	1.72×
	c-trees	0.249	0.422	0.611	Permutation	caudal (contralateral)	3.49×
					Conditional	caudal (contralateral)	3.05×
Quarters	RF	0.306	0.498	0.228	Impurity	middle-caudal (contralateral)	1.29×
					MSE	caudal (contralateral)	1.55×
	c-trees	0.247	0.421	0.614	Permutation	caudal (contralateral)	3.25×
					Conditional	caudal (contralateral)	2.70×
18^{ths}	RF	0.306	0.489	0.276	Impurity	SS04: caudal-anterior	1.47×
					MSE	SS14: middle-posterior	1.42×
	c-trees	0.248	0.420	0.620	Permutation	SS04: caudal-anterior	2.74×
					Conditional	SS04: caudal-anterior	3.85×
96^{ths}	RF	0.302	0.484	0.304	Impurity	SS04: caudal-anterior	2.47×
					MSE	SS26: middle-caudal-anterior	1.78×
	c-trees	0.243	0.417	0.637	Permutation	SS21: caudal-posterior	3.75×
					Conditional	SS04: caudal-anterior	4.04×

important than ipsilateral parotid for segmentation into halves, thirds, and quarters. Therefore, to reduce computational burden, segmentation into 18^{ths} and 96^{ths} used only contralateral parotids. MAE, RMSE, r_{pa}, and summarized importances are shown where applicable. C-tree methods performed significantly better than whole parotid mean dose models and RF (linear and exponential; both $p < 0.0001$) at all segmentation levels. RF methods did not significantly improve prediction when segmentation was introduced ($p \geqslant 0.258$) but c-trees improvement significantly strengthened ($p < 0.039$), improving from a correlation that was already nearly *double* the next-best method (0.531; linear model). Refinement-induced reductions in both MAE and RMSE were similar for RF and c-tree methods (ΔMAE: -0.013 versus -0.016; ΔRMSE: -0.22 versus -0.20). At all levels of segmentation a Kolmogorov–Smirnov test showed no statistically significant differences between the number of supersampled dose matrix voxels contained within each sub-segment ($p > 0.05$ in all $(96{\cdot}95+18{\cdot}17+4{\cdot}3+3{\cdot}2+2{\cdot}1)/2 = 4723$ comparisons).

In almost every importance assessment method, a caudal-most sub-segment was most important. In the two exceptions, the most important sub-segment (middle; between caudal-most and cranial-most sub-segments) was either fully or partially within the caudal 50%-volume. In one of these exceptions, the 18^{ths} segmentation RF-MSE case, the next most important non-middle sub-segment was caudal.

The most important sub-segments, on average over all segmentation methods, had importances 2.4× that of an equivalent sub-segment in a theoretical homogeneous parotid (see table 1). This figure increased when segmentation and methodology was refined: 2.7× when only 18^{ths} and 96^{ths} segmentation was considered, 3.0× when only 96^{ths} segmentation was considered, and 4.0× when only c-tree conditional permutation (the most reliable method) was considered at the finest (96^{ths}) segmentation.

Other than the most important sub-segment, the least important sub-segment, median importances of family-wise groupings based on anatomy (e.g. caudal versus middle versus cranial, or anterior versus posterior), and family-wise percentiles (e.g. 20% and 80%) conveyed similarly the importance of caudal aspects. Supplementary tables showing sub-segment importance are given in the supplementary document. Quantitative information for the most reliable tech-

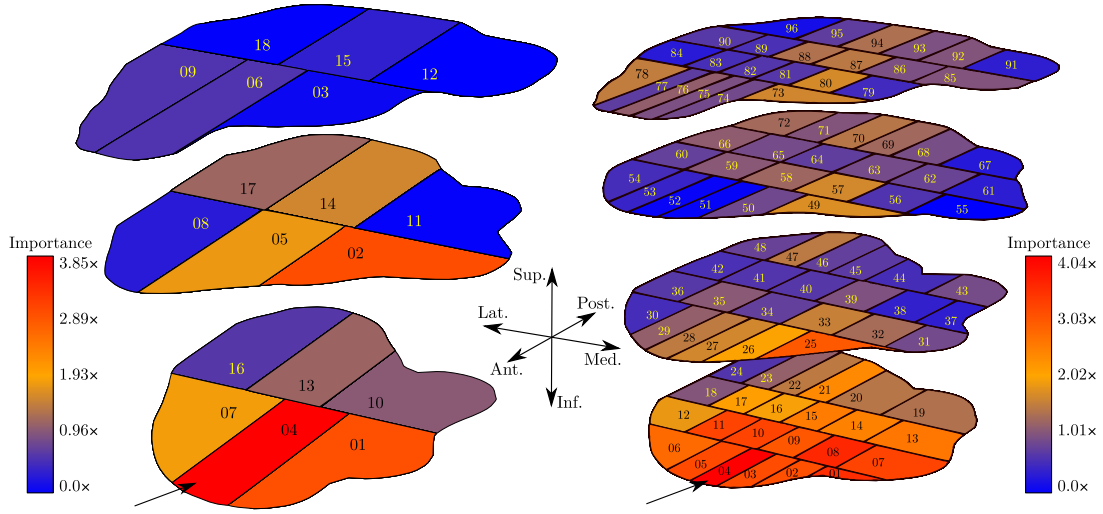

Figure 2. Relative c-tree conditional permutation importance of sub-segments for 18^{ths} (left) and 96^{ths} (right) segmentation. Equal-volume sub-segments are represented by a single slice of axial plane encompassed by the sub-segment. In segmentation into 18^{ths} (96^{ths}), importances span ∼0–3.85× (∼0–4.04×, respectively) that of equivalent sub-segments in a homogeneous parotid. The most important sub-segments are indicated.

nique, c-tree conditional permutation importance, is displayed in the form of heat maps for 18^{ths} and 96^{ths} segmentation in figure 2.

4. Discussion

Effects that deviate from strict parotid functional-spatial homogeneity have been reported, but there is not yet consensus about the criticality of specific sub-volumes in relation to radiotherapy-induced salivary dysfunction. In this work, a regional effect is characterized via a segmentation refinement method. We improve upon existing studies primarily by being systematic in coverage of the parotid: no aspects were *a priori* selected for study and importance of the whole parotid is simultaneously developed.

Four non-parametric methods were used in this work. Though they varied in susceptibility to multicollinearity and other biases, all confirmed the importance of caudal aspects for predicting radiotherapy-induced late salivary function. Contralateral parotids were found to be most important, which is consistent with much of the literature (e.g. [5]). Sub-segment heat maps overlapped across segmentations and importance methods, which suggests conclusions do not substantially depend on the spatial resolution or other convergence factors (e.g. number of trees). A gradient of importance emerged indicating both caudal-anterior aspects are most important and that importance gradually fades posteriorly and superiorly. Starting at the most important sub-segment, movement to superiorly-adjacent regions affected the greatest reduction in importance. Posterior movement less so, and medial and lateral movement affected importance only weakly and approximately equally. Lack of medial-lateral preference may result from parotid medial shrinkage during radiotherapy [29]; lateral aspects may have traveled medially and 'smeared' importance. It remains to be seen if this effect is a treatment artifact.

C-tree methods outperformed RF significantly, and while they both generally improved MAE and RMSE as segmentation proceeded, only the c-tree r_{pa} significantly improved. It is not possible to ascribe this to any specific factor, but it is likely that either (1) RF is intrinsically not capable of ferreting out the information that an equivalent c-tree ensemble can, or (2) RF was strongly impacted by multicollinearity or measurement noise and tree construction was biased. In either case, while RF did not significantly perform *better* than whole parotid models, it was also not significantly *worse*, and we therefore believe it remains a valid tool for inspecting sub-segment importance. C-tree methods, however, outperformed both RF and whole parotid models in every case.

Though there is general consensus among researchers that the parotid is not homogeneous, there is little consensus about the specifics of the inhomogeneities. The existence of critical regions, mechanisms supporting them, and comparative clinical relevance of various aspects and lobes are debated. Though the analysis presented here cannot definitively demonstrate universal clinical relevance, it is constructive to compare to other recent findings. The region we have found to be most important overlaps, at least somewhat, with critical regions reported in previous studies. Buettner *et al* in 2012 compared the relative importance of 50 clinical and physical factors (both categorical and continuous) for subjective xerostomia in 63 head-and-neck cancer patients [8]. Four of the seven most important regressors (mean dose to either parotid, contralateral parotid caudal-medial aspect dose concentration, and contralateral parotid superficial lobe cranio-caudal dose distribution) displayed

agreement with our findings. Regressor importance changed when sub-cohorts were evaluated, but caudal aspects remained important. They concluded, however, that minimizing dose to the lateral and cranial aspects would reduce xerostomia incidence. Our relative importance assessments are in broad agreement, but our conclusions about clinical relevancy differ. Owing to the complexity of head-and-neck anatomy, minimizing dose to lateral and cranial aspects generally requires *increasing* caudal aspect dose. As we collectively have found caudal aspects to be important for clinical outcomes, the recommendation is surprising and implies our interpretation of *prediction* importances and *outcomes* importances differ. In recent work by Clark *et al*, a model-based approach incorporating sensitivity analysis was used to assess relative importance. Linear models performed best and the collective caudal aspect slopes were both most important and largest in magnitude, implying that shifting dose to the caudal aspects would overall negatively impact salivary function. Similar findings have been reported by others [13]. We therefore believe that regressor importance (in this case) translates to clinical relevance. Differences in study designs, outcomes, assessment, cohort size and demographics, and factors considered (especially their response shape) may have contributed to the discrepancy. However, our clinical recommendations are in agreement when the caudal aspects are dose-saturated and cranial or posterior aspects can be spared by shifting dose to the (already saturated) caudal aspects, which may reduce dysfunction. This common clinical situation demonstrates that characterization of regional effects throughout entire parotids can improve outcomes risk analysis compared to simple recommendations to spare specific regions or lobes.

Ortholan *et al* found in 2009 that the contralateral parotid volume receiving $\geqslant$40Gy (V_{40}) was the best dose-volume factor for predicting recovery of salivary function [5]. This finding suggests the non-equivalency of whole mean dose and V_{40}—both of which are dose-volume measures. Deviations from expected dose-volume effects, which follow directly from inhomogeneous radiosensitive structure distribution, have been known for several decades [30, 31]. While the findings of Ortholan *et al* do not specifically describe a regional effect, the regions selected by our two approaches may overlap. Since standard clinical practice involves preventative irradiation of lymph nodes in the head-and-neck, proximate caudal parotid aspects often receive the highest dose. Therefore, V_{40} may simply be selecting the aspects, which would represent a dose-volume manifestation of a regional effect. We believe the reverse (caudal aspect importance reflecting V_{40}) is not true because contralateral parotid (lower dose) was found to be more important than ipsilateral parotid (higher dose), and axially the regions of highest dose follow a medial-anterior to lateral-posterior ridge. Both demonstrate that importance does not merely reflect the dose distribution across the parotid. A more recent report by van Luijk *et al* showed the presence of a confined critical region in the medial-dorsal aspects adjacent to mandible [3]. While our findings are not quite consistent in the superior-inferior direction, they appear to coincide in the anterior-posterior and medial-lateral directions. Both may coincide with major ducts, vasculature, or interfere with innervation; previous real-time imaging of stimulated parotids showed increased perfusion variability focused in the vicinity of both regions [32]. The hypothesis of van Luijk *et al* that damage to stem/progenitor cells is the underlying cause of dysfunction, if true, would support ducts rather than vasculature or nerve impairment. The conclusion of a well-confined critical zone, however, was not confirmed in this work. We found that even very small regions are not necessarily 'critical'. At best, the most important sub-segments appear to have 4$\times$ the importance that a homogeneous parotid sub-segment would. It is possible that population averaging has 'smeared' importance. On the other hand, importance of the most caudal-anterior aspects were, in some cases, two orders of magnitude or greater than cranial and posterior sub-segments and naturally formed smooth importance gradients, which suggests an *effectively* critical (but somewhat broad and smeared) clinically relevant region. A smeared critical region would be more consistent with Lyman normal tissue complication probability models with parallel volume dependence parameters than confined critical regions, and may more accurately reflect stem/progenitor cell distribution [13]. Additional work is needed to characterize this effect.

Both Buettner *et al* and van Luijk *et al* report observing a bath-and-shower effect, which may confound importance assessment, especially for intensity-modulated radiotherapies. Knock-on effects (indeed, also higher-order interactions) are accounted for in RF and c-trees by permutation-based importances [23]. Explicitly including all first-order interaction terms for verification was not feasible even for segmentation into 18$^{\text{ths}}$ owing to the increased complexity and decreased statistical power (i.e. a total of 171 regressors would need to be considered; n.b. $N = 332$). Heterogeneous segmentation could in principle alleviate such issues, but it then becomes unclear how to robustly map regressor importance to clinical relevance.

The uncertainty of our findings is hard to directly quantify. Nested segmentation achieved a precision such that sub-segment volume deviated by less than 0.1%, so the largest source of uncertainty stems from parotid gland delineation (contouring). There is a possibility that large, systematic differences arising from varying clinical practices will result in a set of findings that are internally consistent, but cannot generalize to other institutions. Previous comparisons with other centres suggests both that our institutional contouring practices are reasonably consistent, and that there is a reasonable likelihood of compatibility for derived salivary models [33, 34]. Besides gross errors, it is pos-

sible for small contouring errors to perturb importance analysis. Nested segmentation was developed specifically to mitigate the issue of unreliable contouring by ensuring sub-segments are consistently located in parotids with differing morphologies, as depicted in figure 1. Furthermore, contour quality was scrutinized by a single senior head and neck oncologist. Besides analytical robustness and quality assurance, we believe the emergence of high importance regions that consistently overlap at different scales is strong evidence that our results are meaningful. A noise-dominated analysis would have broad, 'smeared' regions of importance that emerge gradually as resolution increases, or disconnected, randomly fluctuating pockets of importance. Neither were observed; the importances obtained with 18 sub-segments mirror the importances obtained with 96.

In conclusion, caudal-anterior aspects of the parotid were found to be most important for prediction of radiation-induced late baseline-normalized salivary flow. Conditional inference trees, combined with fine segmentation, were found to significantly outperform whole parotid mean dose for prediction of salivary dysfunction.

Acknowledgments

This work supported by the Walter C. Sumner Foundation.

Conflict of interest statement

The authors report no conflict of interests.

ORCID iDs

Haley D Clark https://orcid.org/0000-0003-4780-8093

References

[1] Deasy J O *et al* 2010 Radiotherapy dose–volume effects on salivary gland function *Int. J. Radiat. Oncol. Biol. Phys.* **76** S58–63
[2] Bentzen S M *et al* 2010 Quantitative analyses of normal tissue effects in the clinic (QUANTEC): an introduction to the scientific issues *Int. J. Radiat. Oncol. Biol. Phys.* **76** S3–9
[3] van Luijk P *et al* 2015 Sparing the region of the salivary gland containing stem cells preserves saliva production after radiotherapy for head and neck cancer *Sci Transl. Med.* **7** 305ra147
[4] Clark H *et al* 2015 Regional radiation dose susceptibility within the parotid gland: effects on salivary loss and recovery *Med. Phys.* **42** 2064–71
[5] Ortholan C *et al* 2009 Modeling of salivary production recovery after radiotherapy using mixed models: determination of optimal dose constraint for IMRT planning and construction of convenient tools to predict salivary function *Int. J. Radiat. Oncol. Biol. Phys.* **73** 178–86
[6] Wang Z H *et al* 2011 Impact of salivary gland dosimetry on post-IMRT recovery of saliva output and xerostomia grade for head-and-neck cancer patients treated with or without contralateral submandibular gland sparing: a longitudinal study *Int. J. Radiat. Oncol. Biol. Phys.* **81** 1479–87
[7] van Luijk P *et al* 2009 Bath and shower effects in the rat parotid gland explain increased relative risk of parotid gland dysfunction after intensity-modulated radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **74** 1002–5
[8] Buettner F *et al* 2012 Novel approaches to improve the therapeutic index of head and neck radiotherapy: an analysis of data from the PARSPORT randomised phase III trial *Radiother. Oncol.* **103** 82–7
[9] Miah A *et al* 2016 Recovery of salivary function: contralateral parotid-sparing intensity-modulated radiotherapy versus bilateral superficial lobe parotid-sparing intensity-modulated radiotherapy *Clin. Oncol.* **28** e69–76
[10] Xiao W *et al* 2016 A split-parotid delineation approach for dose optimization in volumetric modulated arc therapy for nasopharyngeal carcinoma patients with parapharyngeal space invasion and level IIa cervical lymph node involvements *Br. J. Radiol.* **89** 20150635
[11] Konings A W T *et al* 2005 Volume effects and region-dependent radiosensitivity of the parotid gland *Int. J. Radiat. Oncol. Biol. Phys.* **62** 1090–5
[12] Zhang H *et al* 2013 Superficial parotid lobe–sparing delineation approach: a better method of dose optimization to protect the parotid gland in intensity-modulated radiotherapy for nasopharyngeal carcinoma *Curr. Oncol.* **20** e577
[13] Roesink J M *et al* 2001 Quantitative dose-volume response analysis of changes in parotid gland function after radiotheraphy in the head-and-neck region *Int. J. Radiat. Oncol. Biol. Phys.* **51** 938–46
[14] Chao K *et al* 2001 A prospective study of salivary function sparing in patients with head-and-neck cancers receiving intensity-modulated or three-dimensional radiation therapy: initial results *Int. J. Radiat. Oncol. Biol. Phys.* **49** 907–16
[15] Clark H *et al* 2014 Automated segmentation and dose-volume analysis with DICOMautomaton *J. Phys.: Conf. Ser.* **489** 012009
[16] Clark H *et al* 2014 Semi-automated contour recognition using DICOMautomaton *J. Phys.: Conf. Ser.* **489** 012088
[17] Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32
[18] Liaw A and Wiener M 2002 Classification and regression by randomForest *R. News* **2** 18–22
[19] Altmann A *et al* 2010 Permutation importance: a corrected feature importance measure *Bioinformatics* **26** 1340–7
[20] Strobl C *et al* 2007 Bias in random forest variable importance measures: illustrations, sources and a solution *BMC Bioinform.* **8** 1
[21] Hothorn T, Hornik K and Zeileis A 2006 Unbiased recursive partitioning: a conditional inference framework *J. Computat. Graph. Stat.* **15** 651–74
[22] Strobl C, Malley J and Tutz G 2009 An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests *Psychol. Methods* **14** 323
[23] Strobl C, Hothorn T and Zeileis A 2009 Party on! a new, conditional variable-importance measure for random forests available in the party package *R. J.* **1** 14–7
[24] Akaike H 1974 A new look at the statistical model identification *IEEE Trans. Autom. Control* **19** 716–23
[25] Stone M 1977 An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion *J. R. Stat. Soc.* B **39** 44–7
[26] Willmott C J and Matsuura K 2005 Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance *Clim. Res.* **30** 79–82
[27] Chai T and Draxler R R 2014 Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature *Geosci. Model Dev.* **7** 1247–50
[28] Dunn O J and Clark V 1971 Comparison of tests of equality of dependent correlation coefficients *J. Am. Stat. Assoc.* **66** 904–8
[29] Nishi T *et al* 2013 Volume and dosimetric changes and initial clinical experience of a two-step adaptive intensity modulated

radiation therapy (IMRT) scheme for head and neck cancer *Radiother. Oncol.* **106** 85–9
[30] Hopewell J W and Trott K R 2000 Volume effects in radiobiology as applied to radiotherapy *Radiother. Oncol.* **56** 283–8
[31] Trott K *et al* 1995 The effect of irradiated volume on the chronic radiation damage of the rat large bowel *Strahlentherapie und Onkologie: Organ der Deutschen Rontgengesellschaft* **171** 326–31
[32] Clark H D *et al* 2015 Development of a method for functional aspect identification in parotid using dynamic contrast-enhanced magnetic resonance imaging and concurrent stimulation *Acta Oncol.* **54** 1686–90
[33] Thor M *et al* 2017 Internal and external generalizability of temporal dose–response relationships for xerostomia following IMRT for head and neck cancer *Radiother. Oncol.* **122** 200–6
[34] Moiseenko V *et al* 2012 Treatment planning constraints to avoid xerostomia in head-and-neck radiotherapy: an independent test of QUANTEC criteria using a prospectively collected dataset *Int. J. Radiat. Oncol. Biol. Phys.* **82** 1108–14