

Journal Pre-proof

Head and Neck Radiotherapy Patterns of Practice Variability Identified as a Challenge to Real-World Big Data: results from the Learning from Analysis of Multicentre Big Data Aggregation (LAMBDA) Consortium

Amanda Caissie MD PhD , Michelle Mierzwa MD , Clifton David Fuller MD PhD , Murali Rajaraman MD , Alex Lin MD , Andrew MacDonald MD , Richard Popple PhD , Ying Xiao PhD , Lianne VanDijk MD , Peter Balter PhD , Helen Fong , Heping Xu PhD , Matthew Kovoov , Joonsang Lee PhD , Arvind Rao PhD , Mary Martel PhD , Reid Thompson MD , Brandon Merz PhD , John Yao PhD , Charles Mayo PhD



PII: S2452-1094(22)00032-X
DOI: <https://doi.org/10.1016/j.adro.2022.100925>
Reference: ADRO 100925

To appear in: *Advances in Radiation Oncology*

Received date: 11 July 2021
Accepted date: 24 December 2021

Please cite this article as: Amanda Caissie MD PhD , Michelle Mierzwa MD , Clifton David Fuller MD PhD , Murali Rajaraman MD , Alex Lin MD , Andrew MacDonald MD , Richard Popple PhD , Ying Xiao PhD , Lianne VanDijk MD , Peter Balter PhD , Helen Fong , Heping Xu PhD , Matthew Kovoov , Joonsang Lee PhD , Arvind Rao PhD , Mary Martel PhD , Reid Thompson MD , Brandon Merz PhD , John Yao PhD , Charles Mayo PhD , Head and Neck Radiotherapy Patterns of Practice Variability Identified as a Challenge to Real-World Big Data: results from the Learning from Analysis of Multicentre Big Data Aggregation (LAMBDA) Consortium, *Advances in Radiation Oncology* (2022), doi: <https://doi.org/10.1016/j.adro.2022.100925>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier Inc. on behalf of American Society for Radiation Oncology.
This is an open access article under the CC BY-NC-ND license
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

**Head and Neck Radiotherapy Patterns of Practice Variability Identified as a Challenge to
Real-World Big Data: results from the Learning from Analysis of Multicentre Big Data
Aggregation (LAMBDA) Consortium**

Authors

Amanda Caissie MD PhD¹, Michelle Mierzwa MD², Clifton David Fuller MD PhD³, Murali Rajaraman MD⁴, Alex Lin MD⁵, Andrew MacDonald MD⁶, Richard Popple PhD⁶, Ying Xiao PhD⁵, Lisanne VanDijk MD³, Peter Balter PhD³, Helen Fong¹, Heping Xu PhD¹, Matthew Kovoov², Joonsang Lee PhD², Arvind Rao PhD², Mary Martel PhD³, Reid Thompson MD⁷, Brandon Merz PhD⁷, John Yao PhD², Charles Mayo PhD²

1 Dalhousie University, Saint John, NB, Canada

2 University of Michigan, Ann Arbor, MI 48109

3 MD Anderson Cancer Center, Houston, TX

4 Dalhousie University, Halifax Nova Scotia, Canada

5 University of Pennsylvania, Philadelphia PA

6 University of Alabama, Birmingham, AL

7 University of Oregon Health Sciences Center, OR

Corresponding Author:

Charles Mayo, PhD

University of Michigan

cmayo@med.umich.edu

734-232-3837

Conflicts of Interest:

Charles Mayo has a research grant from Varian Medical Systems

Ying Xiao has grants from NCI 2U24CA180803-06(IROC), 2U10CA180868-06(NRG)

Andrew MacDonald has a research grant from Varian Medical Systems

Funding: None

Running Title: LAMBDA Consortium

Journal Pre-proof

Abstract

Purpose/Objective: Outside of randomized clinical trials, it is difficult to develop clinically relevant evidence-based recommendations for radiotherapy (RT) practice guidelines due to lack of comprehensive real-world data. To address this knowledge gap, we formed the Learning and Analytics from Multicenter Big Data Aggregation (LAMBDA) consortium to cooperatively implement RT data standardization, develop software solutions for data analysis and recommend clinical practice change based on real-world data analyzed. The first phase of this “Big Data” study aimed at characterizing variability in clinical practice patterns of dosimetric data for organs at risk (OAR), that would undermine subsequent use of large scale, electronically aggregated data to characterize associations with outcomes. Evidence from this study was used as the basis for practical recommendations to improve data quality.

Materials/Methods: Dosimetric details of patients with H&N cancer treated with RT between 2014 and 2019 were analyzed. Institutional patterns of practice were characterized including structure nomenclature, volumes and frequency of contouring. Dose volume histogram (DVH) distributions were characterized and compared to institutional constraints and literature values.

Results: Plans for 4664 patients treated to a mean plan dose of 64.4 ± 13.2 Gy in 32 ± 4 fractions were aggregated. Prior to implementation of TG263 guidelines in each institution, there was variability in OAR nomenclature across institutions and structures. With evidence from this study, we identified a targeted and practical set of recommendations aimed at improving the quality of real-world data.

Conclusion: Quantifying similarities and differences among institutions for OAR

structures and DVH metrics is the launching point for next steps to investigate potential relationships between DVH parameters and patient outcomes.

Journal Pre-proof

Introduction

We know a great deal about the small percentage of patients who are treated on randomized controlled trials but comparatively little about the treatment approaches and outcomes for the large percentage of patients treated in routine practice. While clinical trials are thought to be the optimal method to demonstrate causal effects between treatments and outcomes, results may not generalize to the majority of patients who are treated in the real-world setting.¹ There is need for comprehensive real-world data characterizing efficacy and toxicity of anti-cancer treatments such as radiotherapy (RT).² Large scale, real-world data has potential not only to augment clinical trial design and validation³ but also to improve RT plan quality and patient outcomes by characterizing and reducing practice variability.⁴ However, assessment of RT clinical practice patterns is challenged by the complexity of non-standardized electronic dosimetric data. International efforts are currently underway to promote standardization of RT data, from the American Society of Radiation Oncology (ASTRO)'s consensus papers recommending standardized RT normal tissue contouring, minimum data sets (MDS) and synoptic treatment summaries, to the American Association of Physicists in Medicine (AAPM)'s TG-263 guideline for standardized RT nomenclature and the Canadian Partnership for Quality Radiotherapy (CPQR)'s ongoing development of Patient Reported Outcome and Big Data guidance documents. These RT society led quality improvement (QI) initiatives are helping to pave the way toward facilitated capture and use of RT "Big Data".⁵⁻⁹

Presently, little is known about the quality of existing RT data because variabilities in clinical practice challenge automated data pooling and resource exhaustive manual approaches do not scale. Large scale, multi-institutional dosimetric data could be highly valuable for clinical assessment of treatment plan quality and for modeling associations with treatment toxicities. As the first of its kind to assess head and neck (H&N) RT “Big Data”, we have formed the Learning and Analytics from Multicenter Big Data Aggregation (LAMBDA) consortium to cooperatively implement RT data standardizations, develop software solutions for data aggregation, and recommend clinical practice changes based on real-world data analysis. The first phase of this “Big Data” study across 5 international institutions aimed to compare organ at risk (OAR) nomenclature, dose volume histogram (DVH) metric norms and institutional patterns of practice across large numbers of patients treated with RT for H&N cancer. Our results from the amalgamated multi-institutional data have been used to develop evidenced based RT plan quality recommendations for structures and dose volume histogram metrics for LAMBDA members. The aim of these recommendation is to support interoperable data exchange and pooling, by reducing variability in clinical practice to facilitate learning from large scale standardized real-world dosimetric treatment data.

Materials/Methods

Six institutions are currently participants in the LAMBDA consortium: xx,xx,xx,xx,xx,xx. Five have completed DVH data submission. Operating under a Research Ethics Board approved protocol, the group aggregated retrospective data of H&N cancer RT plans from 2014 and 2019. All LAMBDA institutions have now implemented TG-263 guidelines for standardized RT

nomenclature, with 2019 being the year that such standardization was finalized. All datasets were anonymized by the submitting institutions before submission to the central institution (xxxxx) for aggregation and analysis. The treatment planning systems used were Varian Eclipse 13.6 and 11.5 or Pinnacle 16.2. Custom applications for automated extraction and aggregation into a database (Microsoft MS-SQL v 12.0) were used. All analyses were carried out using R (v 3.4.1), a computational statistics software package. Significance of differences in comparison of any values was determined with the student's-t test using a threshold of $p=0.05$.

Structure names for RT plans completed before TG-263 guideline implementation were resolved by mapping OAR names to TG-263 standard values. Variability in OAR nomenclature was measured by counting alternative mappings for each OAR name and institution. For each plan, if both left and right parallel function structures were drawn (e.g. Parotid_L and Parotid_R), they were additionally sub- categorized according to relative mean dose (e.g. Parotid_High and Parotid Low).

Statistical DVH curves were used to visualize quantified comparisons of DVH curves among institutions. This provides a graphical means of identifying inter-institutional variation in practice norms for dose distributions of structures. To detail the frequency of values for each institution, histograms were created for the distribution of structure volumes and DVH metrics. Differences between histograms of distributions for pairs of institutions were calculated by summing $\frac{1}{2}$ of the difference in each histogram bin to calculate the cumulative histogram difference (CHD). For example, if the distribution of structure volume values for two institutions

were identical, then $CHD = 0$. If there was no overlap in the distribution of volumes for the two institutions, then $CHD = 1$. CHD values for each unique pair of institutions were averaged ($\langle CHD \rangle$) to quantify the inter-institutional variability in distributions of structure volume or DVH metrics. In other words, if there were differences among institutions in how they contoured structures, with some contouring them larger and others contouring them smaller, then inter-institutional variability (ie $\langle CHD \rangle$) could be calculated. $\langle CHD \rangle$ values were used to group structure volumes and DVH metrics according to low (<0.4), moderate ($0.4-0.8$) and high (>0.8) inter-institutional variability.

Distributions were further summarized by calculating median, 1st quantile (25% of values $\leq Q1$) and 3rd quantile (75% of values $\leq Q3$) values. Intra-Institutional variability was quantified by averaging values for $(Q3-Q1)/\text{median}$ (i.e. $\langle (Q3-Q1)/\text{median} \rangle$). A k-means clustering algorithm was used as an objective means to group structures by $\langle (Q3-Q1)/\text{median} \rangle$ and $\langle CHD \rangle$ values, providing a visualization of structure volumes according to the amount of intra and inter-institution variability.

RT planning DVH metrics used by institutions were summarized and compared to the wide range of metrics seen in the literature. For structures with high dose constraints, the $D_{0.03cc}[Gy]$ was calculated and compared across institutions. “Real World” treated values were represented as Median [Q1,Q3] and inter-institutional variability in these dose values was assessed using $\langle CHD \rangle$. When only a portion of the structure volume is drawn (e.g. entire esophagus VS a portion of esophagus proximal to the target), DVH metrics based on a percentage of the structure

(e.g. V50Gy[%]) may be less consistent than metrics based on the absolute volume (e.g. V50Gy[cc]). Absolute volume (VxGy[cc]) CHD were compared to percentage volume versions (VxGy[%]), using <CHD> values to confirm whether the absolute volume DVH metric had less inter-institutional variability. Owing to the proximity of structures, DVH metric values for one structure may be predictive of DVH metric values in other structures. To understand which set of structures have strong dosimetric associations, we used an unsupervised learning approach with Bayesian networks (bnlearn v 4.5).

Consensus Recommendation

Results of the multi-institutional quantitative OAR metric comparisons were reviewed in the context of RT planning constraints set by individual LAMBDA institutions and DVH metric recommendations from the literature. A set of practical H&N RT plan quality recommendations were then developed for LAMBDA members to reduce inter-institutional variability.

Results

Data was analyzed from 4,664 patients from 5 LAMBDA institutions. Average age of patients was 60.1 ± 11 years. Categorized by PTV dose, the cohort was made up of $2 \pm 2\%$ palliative (≤ 50 Gy), $41 \pm 7\%$ adjuvant (>50 Gy and < 70 Gy) and $56 \pm 8\%$ definitive (≥ 70 Gy) cases. For RT of curative or adjuvant intent in a variety of H&N malignancies, high dose PTVs were treated to a mean total dose of 64.4 ± 13.2 Gy in 32 ± 4 fractions. Most institutions (3/5) did not have systems in place to automate electronic extraction of staging information over the years of the study. Of

the two institutions able to report staging information their distributions were stage I (4%,6%), II (5 %, 9%), III (26%,18%), IV (65 %,67%).

Variability of OAR Nomenclature and Contour Inclusion

Figure 1 summarizes OAR structures included in the H&N RT plan data sets. Prior to implementation of TG263 guidelines in each institution, there was variability in OAR nomenclature across institutions. For example, Parotid_L, Left Parotid and Lt Parotid are 3 name variants for Parotid_L. Institution A had the lowest number of name variants per structure (1.5 mean \pm 0.8 standard deviation), followed by institutions C (2.0 \pm 1.3), D (3.2 \pm 2.4), B (3.4 \pm 2.5) and E (6.3 \pm 8.4). Bilateral OAR structures showed substantial variability in nomenclature (5.9 \pm 4.9): optic nerves (4.4 \pm 2.9), lacrimal (3.5 \pm 2.4), parotid (6.3 \pm 5.7) and submandibular glands (11.6 \pm 13.5), with the latter structure being the one with the most name variants across institutions. Institutional guidelines for which structures to routinely contour ranged from a comprehensive set of structures for all patients to a minimal standard set of structures with additional structures contoured only if at risk and potentially spared in treatment planning. Some institutions reported shifting over time from the minimal approach toward a more comprehensive standard set of structures to be contoured.

Of the 54 OAR structures used by the institutions only 2 , SpinalCord and at least one Parotid, were contoured for $\geq 90\%$ of all patients treated. Six structures were contoured for $\geq 90\%$ of patients at 3/5 of institutions: SpinalCord, Brainstem, at least one Parotid, Bone_Mandible,

Larynx, and Esophagus. Lowering the threshold to 50% of patients treated at 3/5 institutions, an additional 7 were identified: oral cavity, brain and at least one of bilateral structures: submandibular glands, eyes, cochleas, optic nerves and a structure to monitor dose to constrictor muscles (e.g Pharynx, *Musc_Constrict_S*, *Musc_Constrict_I*, *Musc_Constrict-PTV*). For planning organ at risk volumes (PRV), only *SpinalCord_PRV* was routinely ($\geq 75\%$ of patients) contoured in the majority ($\geq 3/5$) of institutions. Both parotids were contoured in at least 84.3% of plans (institution A), while the other institutions had higher rates of contouring the bilateral structure (89.6%, 87.9%, 96.6% and 89.8% for institutions B-E respectively). Both submandibular glands were contoured in only 22.7% of plans in one institution (E) while the other institutions had slightly higher rates of contour inclusion (63.3%, 77.4%, 41.6% and 59.5% for institutions A-D respectively). Institution A routinely ($>90\%$) contoured both superior and inferior constrictor muscles, D contoured *Musc_Constrict-PTV* in 95% of cases, C and B contoured pharynx on 9% and 64% of cases respectively. E contoured constrictors on only 2.4% of cases. Less than 20% of total RT plans included contouring of both parotid glands (Parotids), both submandibular glands (*Gnd_Submands*) and the muscle constrictors (*Musc_Constrict*).

Variability of OAR Volumes

Not only was there substantial variability among institutions in terms of which structures were routinely contoured but also what portion of structures were contoured. Many institutions contoured the cochlea as a whole structure (87.9%, 70.4%, 7.5%, 58% at A, C-E respectively). On the other hand, institution B contoured the cochlea in only 15.4% of cases, while contouring the middle ear and inner ear separately for 65% of RT plans. Contouring of the brain (82.6%, 64.3%, 94.6%, 25%, 30.9% at A-E) or temporal lobe (15%, 0%, 2.4%, 79.0%, 0.2% at A-E)

varied substantially. One institution routinely contoured only the sub-volume of OARs outside of the PTV volume (e.g. Bone_Mandible-PTV). Because they exclude a portion of OAR structures (within the PTV), these partially contoured structures are not dosimetrically interchangeable with complete OARs used to measure associations with toxicities.

Figure 2 summarizes analysis of institutional norms for OAR structure volumes. Among the 13 OAR volumes segmented on $\geq 50\%$ of patients for the majority of institutions, six (Brainstem, Eye, Bone_Mandible, Parotid, submandibular glands (GlnD_Submands) and SpinalCord) had low inter-institution variability ($\langle \text{CHD} \rangle < 0.4$ $p < 0.05$). Volumes for OpticNrv, Esophagus, Cochlea, Cavity_Oral and Larynx had moderate inter-institution variability. Volumes for constrictor muscles had the highest inter-institution variability (> 0.8 $\langle \text{CHD} \rangle$).

Variability of OAR DVH Metrics and Constraints

Substantial variability was seen among institutions in OAR DVH metric-constraints used for RT planning (Table 1). Among 29 structures, 58 distinct metric constraints were identified from institutional templates for RT planning. Of these, only 2 constraints (Parotid:Mean[Gy] and Cavity_Oral:Mean[Gy]) were used for $\geq 3/5$ institutions. No more than 2 institutions agreed on high dose constraints for SpinalCord, Brainstem, or Bone_Mandible. More than 3 different constraints were used by institutions for Mean[Gy] for Esophagus, Larynx, and GlnD_Submand. With the exception of institution B, all other institutions prioritized meeting OAR constraints as more (1) or less (3) important with respect to target coverage (2). Amongst the institutions that prioritized high dose constraints for critical structures of SpinalCord, Brainstem and OpticNrvs

and OpticChiasm, priority 1 was assigned for these structures. There was inter-institutional variability in inclusion of priority levels for DVH metrics of other structures and critical structures of SpinalCord_PRV (3/4 institutions), OpticNrv/OpticChiasm (2/4 institutions), and Brainstem PRV (1/4 institutions). There were unique patterns of practice identified for individual institutions such as institution A that assigned priority 1 to Larynx and Musc_Constrict_I, while other institutions either did not assign priority level to the planning constraints of those structures or assigned a level 3.

Real-world treated values were substantially less than institutional or literature-based constraints for some structures. Q3 values were significantly lower than priority 1 constraints for SpinalCord: D0.03cc[Gy] \leq 45, Brainstem:D0.03cc[Gy] \leq 54 and OpticNrv/OpticChiasm:D0.03cc[Gy] $<$ 54. For priority 3 constraints, Q3s were significantly lower than Larynx:Mean[Gy] \leq 45, Cochlea:Mean[Gy] \leq 30 and Esophagus:Mean[Gy] \leq 45. Median values for priority 3 constraints were regularly exceeded for certain OAR (e.g. GlnD_Submand_Low:Mean[Gy]). There may be variability in what fraction of a structure is contoured. For example, what is designated as “esophagus” may correspond to the entire length or only a portion of it proximal to the target volume. Absolute volume (VxGy[cc]) metrics did show low inter-institutional variability (<0.4 <CHD>) compared to percentage volume versions (VxGy[%]) that showed moderate inter-institution variability (0.4-0.8 <CHD>) for the structures of Esophagus, Larynx and Parotid. This highlighted potential for use of partial volume DVH metrics based on absolute vs percentage volumes as a means to mitigate impact of contour practice variability for these structures.

Contour variability is but one factor contributing to RT plan variability, as evidenced by statistical DVH curves (Figure 3) showing dose variations not only for a structure of moderate volume variability (larynx) but also for GlnD_Submand_Low, a structure with low inter-institutional volume variability. Histogram analysis provides visualization of the intra and inter-institutional variability for a single DVH metric (Mean) for structures such as GlnD_Submand_Low or Larynx (Figure 4). Assessment of only an average Median value or quantiles ($\langle \text{Median [Q1, Q3]} \rangle$) of one metric such as Mean (Table 1: GlnD_Submand_Low 48Gy[34,59] and Larynx 33Gy[27,41]), may lead to under-appreciation of the wide range of doses accepted by institutions. Although causal relationships cannot be drawn from the data currently available in this phase of the study, it is noted that institution A accepted a much narrower range of dose for larynx Mean and gave this structure's dose constraints a priority 1 for RT planning whereas other institutions had assigned it a priority 3.

Bayesian network analysis (Figure 5) identified strong predictive associations among DVH metrics of various structures, with GlnD_Submand_Low (≥ 30 Gy) having the largest number of relationships with other structures. Plotting the strength of interactions among structure-DVH metrics was used to identify structures to consider in recommendations for a minimum contour set. The relationships between structures highlights the importance of complete contour sets to facilitate creating multi-structure models of toxicity.

Current Recommendations

From the evidence of this study's results, the consortium identified clinical practice

recommendations for LAMBDA members to support interoperable data exchange and pooling, by reducing variability in clinical practice to facilitate learning from large scale standardized real-world dosimetric treatment data.

- Implement routine and standardized collection of data such as diagnosis and staging in formats that can be easily extracted from electronic systems.
- Adopt TG-263 nomenclature for all OARs and converge on a minimal set of TG-263 compliant target (PTV, CTV, GTV) names acceptable at each institution.
- As a means of ensuring complete datasets, include the 13 structures contoured on $\geq 50\%$ of patients in the majority ($\geq 3/5$) of institutions: brain(Brain), brainstem (Brainstem) , spinal cord (SpinalCord), eyes (Eye_L, Eye_R) , cochleas (Cochlea_L, Cochlea_R) , optic nerve structures (OpticNrv_L, OpticNrv_R, OpticChiasm), mandible (Bone_Mandible), parotids (Parotid_L, Parotid_R) and submandibular glands (GlnD_Submand_R, GlnD_Submand_L), oral cavity (Cavity_Oral) , esophagus (Esophagus) , larynx (Larynx) , and constrictor muscles (Musc_Constric_I, Musc_Constric_S, or Pharynx) for all patients. At minimum, contour those that are within 3 cm of the PTVs. As per ASTRO's recent consensus paper,⁵ other OAR structures should be included based on disease site treated and all OARs should be contoured following published atlases.
- In data pooling applications provide at least the minimum set of 18 DVH metrics for reporting were identified for these 13 structures (● in Table 1).
- Consistent with guidelines, critical structures of SpinalCord_PRV, Brainstem_PRV and

OpticNrvs and OpticChiasm should be assigned priority 1 in RT planning.

- For bilateral, parallel function structures (Parotid_L, Parotid_R, GlnD_Submand_L, GlnD_Submand_R) include both left and right structures if present (i.e. unresected)
- If applicable, contour the larger and more inclusive OAR structures of Brain versus Lobe_Temporal and Cochlea versus division of Ear Middle and Ear_Inner.
- Given reported relationships between dysphagia and dose to individual muscle constrictor components, separation of Musc_Constrict_S and Musc_Constrict_I is recommended versus Pharynx.^{24,27,32}
- If using OAR-PTV volumes, contour the corresponding OAR volume. For high dose values, D0.03cc[Gy] is recommended for data pooling (versus Max[Gy] or D0.1cc[Gy]) to ensure interoperability and consistency with recently published consensus guidelines.⁴¹
- Consider reducing constraint values for DVH metrics where median and Q3 values (Table 1) are well below standard limits set in the literature (e.g., Esophagus).

Discussion

To our knowledge, the present study is the first to use a large combined dataset (> 4000 patients) drawn from “real world” H&N cancer RT practice, to quantify inter-institutional variability in OARs routinely segmented and norms for OAR DVH metric values. The results of the present study are timely and critical given the suite of recent recommendations released from large international organizations, such as ASTRO, AAPM and CPQR, aiming to decrease variability in patterns of practice and promote standardization of RT. The collaborators of this LAMBDA consortium are also actively involved in AAPM and ASTRO’s combined effort to develop an

operational ontology for radiation oncology, that includes professional society endorsed standardizations like TG-263. As such, this research effort aims to identify gaps in clinical practice that need to be addressed in order for ontologies to be successfully applied in routine use. Just a few years ago, a similar consortium in Europe, the ENT COBRA (Consortium for Brachytherapy Data Analysis), paved the way in H&N ontology work to standardize data collection for H&N cancer patients treated with brachytherapy.

Given that OAR structures used in the current analysis are named according to TG-263, this study serves as a means of knowledge translation to promote uptake of AAPM's recommendations which have established a foundation for sharing of large-scale aggregated data without the prohibitive effort of manually extracting variable data. A few publications have emerged with focus on technical approaches to either efficiently re-label retrospective data or improve TG-263 compliance going forward.^{56,57} The current study details the significant variability in OAR nomenclature that was present prior to TG-263 implementation, with certain structures such as submandibular glands having over 35 name variants across institutions.

Significant inter-institutional variability was also found for which structures were included in RT plans. The LAMBDA consortium has therefore recommended that a minimum OAR data set include the 13 structures consistently contoured across the majority of institutions, with other OAR structures to be included based on disease site treated as per recent ASTRO guidelines.⁵ In the present study, less than 20% of total RT plans included a complete set of contours for parotid glands, submandibular glands and muscle constrictors. Strong relationships were identified between DVH metrics of various structures such as GInd_Submand_Low, highlighting the

critical importance of establishing standardized and complete OAR sets if future studies are to be successful in investigation of associations between real-world DVH data and RT toxicities such as xerostomia or dysphagia.

OAR volume variability across institutions was another study finding. It is acknowledged that the present study assessed DVH metrics alone and did not address case-by-case contour variability or OAR contouring practices of individual institutions. This work is considered a logical next step for the LAMBDA consortium, aiming to further standardize OAR contouring based on atlases such as those from the RTOG. As preliminary data, the current study's results allow for inferences to be made with respect to OAR contour patterns of practice based on volume of OARs determined from DVH metrics. While it is inferred that practice variability exists for contours of certain structures (e.g. larynx), it is also hypothesized that there are factors beyond contour variation that may affect dose constraints achieved, given that significant inter-institutional variability was found for DVH metrics of OARs that had similar (e.g. submandibular glands) or dissimilar (e.g. larynx) volumes across institutions.

These results are consistent with a recent report of the 15 Dutch radiation oncology institutions showing large inter-institutional variations in PTV and OAR dosimetry of a benchmarking test case of 1 H&N RT plan with 6 OAR structures.⁵⁹ While OAR sparing improved through collaborative iterations of contour and plan comparisons, unexplained inter-institutional differences still existed across OAR doses despite more consistent contouring. Work is required to investigate the source of such variations, which could include RT planning prioritization of

OAR constraints as more or less important with respect to target coverage. The current study showed inter-institutional variability of such OAR prioritization for critical structures such as spinal cord PRV and brainstem PRV as well as structures such as larynx. Even if OAR constraints are being prioritized, there may be question as to whether they are feasible to achieve if attempted.

Identifying baseline norms from clinical practice enables benchmarks to be set based on routinely achievable values and future avoidance of atypical values. Statistical DVH evaluation of the current study showed that the majority of RT plans achieved dose constraints for OAR such as esophagus well below limits set by LAMBDA institutions and published recommendations such as QUANTEC. With the guiding principle of as low as reasonably achievable (ALARA), institutions may choose to set optimization constraints based on what is achievable in the majority of their own cases or results of such multi-institutional collaborations may allow institutions to strive for plan optimization based on what other institutions have shown to be achievable. On the other hand, this study found that recommended values for certain structures such as Gnd_Submand_Low were often exceeded, highlighting the need to systematically collect patient outcome data in routine practice so that toxicity profiles such as xerostomia may be assessed in future. Evaluations of published models of toxicity, such as QUANTEC, are limited without sufficient real-world data to place recommendations in the context of clinical norms for what is achievable in practice. Characterizing variability in practice norms could improve understanding when real-world patient outcomes do not mirror clinical trial results. To go beyond description of DVH metric practice patterns to recommend standardized DVH metric constraints, anticipated next steps include control of OAR contour variability and

use of real-world data from LAMBDA institutions to analyze DVH metric associations with patient outcomes.

While Big Data efforts have numerous advantages, there are limitations which must be acknowledged. Presently, little is known about the quality of existing “real world” RT data. For elements that are primarily captured as free-text or not routinely captured in electronic systems, resource exhaustive manual approaches for retrospective extraction can be a substantial barrier for large scale real-world data sets. In this study, only 2 of 5 institutions had the human or financial resources to manually extract data for diagnosis and staging. It is expected that AAPM’s soon to be released oncology ontology will help inform all institutions regarding standardized data capture of critical data elements beyond dosimetric data that is already routinely gathered in electronic systems. This should make assembling large scale dosimetric data sets from routine practice more plausible, so long as variabilities in clinical practice can be limited to facilitate analysis of the automated data extracted. Results are often hypothesis generating and lead to more questions. If there is no consensus for dose constraint on an OAR such as GlnD_Thyroid, should the OAR be routinely contoured for information purposes given the challenges of resource constraints? While the present study identified general practice patterns from amalgamated data of all H&N RT plans, it is recognized that future work is required to investigate unique patterns of practice based on treatment intent or H&N cancer subtype. The current study’s results are drawn from participating institutions, which may have unique patterns of practice. As more institutions collaborate, the risk of such bias decreases.

This multi-institutional Big Data study has identified patterns of H&N RT practice variation. Results of this study have shaped H&N RT plan quality recommendations for LAMBDA consortium members to reduce inter-institutional variability that could introduce hidden biases in the interpretation of pooled large-scale, real-world DVH data. Important next steps have been identified to improve plan quality through standardization and facilitate future studies of dosimetric OAR data and patient outcomes. Whereas clinical trial results have shown plan quality variability to negatively impact patient survival⁴, future Big Data studies must investigate whether dosimetric constraint achievement correlates with decreased toxicity and improved quality of life without compromise of oncologic outcomes.

References

1. Franklin JM, Schneeweiss S. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials? *Clin Pharmacol Ther.* 2017;102(6):924-933.
doi:10.1002/cpt.857
2. Use of Electronic Health Record Data in Clinical Investigations Guidance for Industry. Published online 2018:15.
3. Mayo CS, Matuszak MM, Schipper MJ, Jolly S, Hayman JA, Ten Haken RK. Big Data in Designing Clinical Trials: Opportunities and Challenges. *Front Oncol.* 2017;7:187.
doi:10.3389/fonc.2017.00187
4. Peters LJ, O'Sullivan B, Giralt J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02. *J Clin Oncol.* 2010;28(18):2996-3001. doi:10.1200/JCO.2009.27.4498
5. Wright JL, Yom SS, Awan MJ, et al. Standardizing Normal Tissue Contouring for Radiation Therapy Treatment Planning: An ASTRO Consensus Paper. *Pract Radiat Oncol.* 2019;9(2):65-72. doi:10.1016/j.prro.2018.12.003
6. Grant SR, Smith BD, Mayo CS. The Charge to Liberate Siloed Radiation Oncology Treatment Data Through Uniform and Structured Documentation: A Commentary on American Society for Radiation Oncology and Commission on Cancer Recommendations. *Pract Radiat Oncol.* 2020;10(5):304-307. doi:10.1016/j.prro.2019.11.009
7. Hayman JA, Dekker A, Feng M, et al. Minimum Data Elements for Radiation Oncology:

An American Society for Radiation Oncology Consensus Paper. *Pract Radiat Oncol*.

2019;9(6):395-401. doi:10.1016/j.prro.2019.07.017

8. Christodouleas JP, Anderson N, Gabriel P, et al. A Multidisciplinary Consensus Recommendation on a Synoptic Radiation Treatment Summary: A Commission on Cancer Workgroup Report. *Pract Radiat Oncol*. 2020;10(6):389-401.
doi:10.1016/j.prro.2020.01.002
9. Mayo CS, Moran JM, Bosch W, et al. American Association of Physicists in Medicine Task Group 263: Standardizing Nomenclatures in Radiation Oncology. *Int J Radiat Oncol Biol Phys*. 2018;100(4):1057-1066. doi:10.1016/j.ijrobp.2017.12.013
10. Jackson A, Marks LB, Bentzen SM, et al. The lessons of QUANTEC: recommendations for reporting and gathering data on dose-volume dependencies of treatment outcome. *Int J Radiat Oncol Biol Phys*. 2010;76(3 Suppl):S155-160. doi:10.1016/j.ijrobp.2009.08.074
11. Mayo CS, Yao J, Eisbruch A, et al. Incorporating big data into treatment plan evaluation: Development of statistical DVH metrics and visualization dashboards. *Adv Radiat Oncol*. 2017;2(3):503-514. doi:10.1016/j.adro.2017.04.005
12. Marks LB, Yorke ED, Jackson A, et al. Use of normal tissue complication probability models in the clinic. *Int J Radiat Oncol Biol Phys*. 2010;76(3 Suppl):S10-19.
doi:10.1016/j.ijrobp.2009.07.1754
13. Mayo C, Yorke E, Merchant TE. Radiation associated brainstem injury. *Int J Radiat Oncol Biol Phys*. 2010;76(3 Suppl):S36-41. doi:10.1016/j.ijrobp.2009.08.078

14. Kirkpatrick JP, van der Kogel AJ, Schultheiss TE. Radiation dose-volume effects in the spinal cord. *Int J Radiat Oncol Biol Phys.* 2010;76(3 Suppl):S42-49.
doi:10.1016/j.ijrobp.2009.04.095
15. Deasy JO, Moiseenko V, Marks L, Chao KSC, Nam J, Eisbruch A. Radiotherapy dose-volume effects on salivary gland function. *Int J Radiat Oncol Biol Phys.* 2010;76(3 Suppl):S58-63. doi:10.1016/j.ijrobp.2009.06.090
16. Rancati T, Schwarz M, Allen AM, et al. Radiation dose-volume effects in the larynx and pharynx. *Int J Radiat Oncol Biol Phys.* 2010;76(3 Suppl):S64-69.
doi:10.1016/j.ijrobp.2009.03.079
17. Lawrence YR, Li XA, el Naqa I, et al. Radiation dose-volume effects in the brain. *Int J Radiat Oncol Biol Phys.* 2010;76(3 Suppl):S20-27. doi:10.1016/j.ijrobp.2009.02.091
18. Werner-Wasik M, Yorke E, Deasy J, Nam J, Marks LB. Radiation dose-volume effects in the esophagus. *Int J Radiat Oncol Biol Phys.* 2010;76(3 Suppl):S86-93.
doi:10.1016/j.ijrobp.2009.05.070
19. Mayo C, Martel MK, Marks LB, Flickinger J, Nam J, Kirkpatrick J. Radiation dose-volume effects of optic nerves and chiasm. *Int J Radiat Oncol Biol Phys.* 2010;76(3 Suppl):S28-35.
doi:10.1016/j.ijrobp.2009.07.1753
20. Bhandare N, Jackson A, Eisbruch A, et al. Radiation therapy and hearing loss. *Int J Radiat Oncol Biol Phys.* 2010;76(3 Suppl):S50-57. doi:10.1016/j.ijrobp.2009.04.096
21. Li YC, Chen FP, Zhou GQ, et al. Incidence and dosimetric parameters for brainstem

necrosis following intensity modulated radiation therapy in nasopharyngeal carcinoma.

Oral Oncol. 2017;73:97-104. doi:10.1016/j.oraloncology.2017.08.011

22. Yao CY, Zhou GR, Wang LJ, et al. A retrospective dosimetry study of intensity-modulated radiotherapy for nasopharyngeal carcinoma: radiation-induced brainstem injury and dose-volume analysis. *Radiat Oncol.* 2018;13(1):194. doi:10.1186/s13014-018-1105-z
23. Mortensen HR, Jensen K, Aksglæde K, Behrens M, Grau C. Late dysphagia after IMRT for head and neck cancer and correlation with dose-volume parameters. *Radiother Oncol.* 2013;107(3):288-294. doi:10.1016/j.radonc.2013.06.001
24. Eisbruch A, Kim HM, Feng FY, et al. Chemo-IMRT of oropharyngeal cancer aiming to reduce dysphagia: swallowing organs late complication probabilities and dosimetric correlates. *Int J Radiat Oncol Biol Phys.* 2011;81(3):e93-99. doi:10.1016/j.ijrobp.2010.12.067
25. Eisbruch A, Schwartz M, Rasch C, et al. Dysphagia and aspiration after chemoradiotherapy for head-and-neck cancer: which anatomic structures are affected and can they be spared by IMRT? *Int J Radiat Oncol Biol Phys.* 2004;60(5):1425-1439. doi:10.1016/j.ijrobp.2004.05.050
26. Eisbruch A, Lyden T, Bradford CR, et al. Objective assessment of swallowing dysfunction and aspiration after radiation concurrent with chemotherapy for head-and-neck cancer. *Int J Radiat Oncol Biol Phys.* 2002;53(1):23-28. doi:10.1016/s0360-3016(02)02712-8
27. Caudell JJ, Schaner PE, Desmond RA, Meredith RF, Spencer SA, Bonner JA. Dosimetric factors associated with long-term dysphagia after definitive radiotherapy for squamous cell

- carcinoma of the head and neck. *Int J Radiat Oncol Biol Phys.* 2010;76(2):403-409.
doi:10.1016/j.ijrobp.2009.02.017
28. Ozkaya Akagunduz O, Eyigor S, Kirakli E, et al. Radiation-Associated Chronic Dysphagia Assessment by Flexible Endoscopic Evaluation of Swallowing (FEES) in Head and Neck Cancer Patients: Swallowing-Related Structures and Radiation Dose-Volume Effect. *Ann Otol Rhinol Laryngol.* 2019;128(2):73-84. doi:10.1177/0003489418804260
29. Kamal M, Mohamed ASR, Volpe S, et al. Radiotherapy dose-volume parameters predict videofluoroscopy-detected dysphagia per DIGEST after IMRT for oropharyngeal cancer: Results of a prospective registry. *Radiother Oncol.* 2018;128(3):442-451.
doi:10.1016/j.radonc.2018.06.013
30. Feng FY, Kim HM, Lyden TH, et al. Intensity-modulated radiotherapy of head and neck cancer aiming to reduce dysphagia: early dose-effect relationships for the swallowing structures. *Int J Radiat Oncol Biol Phys.* 2007;68(5):1289-1298.
doi:10.1016/j.ijrobp.2007.02.049
31. Dean J, Wong K, Gay H, et al. Incorporating spatial dose metrics in machine learning-based normal tissue complication probability (NTCP) models of severe acute dysphagia resulting from head and neck radiotherapy. *Clin Transl Radiat Oncol.* 2018;8:27-39.
doi:10.1016/j.ctro.2017.11.009
32. Chera BS, Fried D, Price A, et al. Dosimetric Predictors of Patient-Reported Xerostomia and Dysphagia With Deintensified Chemoradiation Therapy for HPV-Associated Oropharyngeal Squamous Cell Carcinoma. *Int J Radiat Oncol Biol Phys.* 2017;98(5):1022-

1027. doi:10.1016/j.ijrobp.2017.03.034
33. 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(2):127-136. doi:10.1016/S1470-2045(10)70290-4
34. Murdoch-Kinch CA, Kim HM, Vineberg KA, Ship JA, Eisbruch A. Dose-effect relationships for the submandibular salivary glands and implications for their sparing by intensity modulated radiotherapy. *Int J Radiat Oncol Biol Phys.* 2008;72(2):373-382. doi:10.1016/j.ijrobp.2007.12.033
35. Jackson WC, Hawkins PG, Arnould GS, Yao J, Mayo C, Mierzwa M. Submandibular gland sparing when irradiating neck level IB in the treatment of oral squamous cell carcinoma. *Med Dosim.* 2019;44(2):144-149. doi:10.1016/j.meddos.2018.04.003
36. Hawkins PG, Lee JY, Mao Y, et al. Sparing all salivary glands with IMRT for head and neck cancer: Longitudinal study of patient-reported xerostomia and head-and-neck quality of life. *Radiother Oncol.* 2018;126(1):68-74. doi:10.1016/j.radonc.2017.08.002
37. Chen AM, Yoshizaki T, Wang PC, et al. Hazards of sparing the ipsilateral parotid gland in the node-positive neck with intensity modulated radiation therapy: Spatial analysis of regional recurrence risk. *Adv Radiat Oncol.* 2018;3(2):111-120. doi:10.1016/j.adro.2017.12.004
38. Huang J, He T, Yang R, Ji T, Li G. Clinical, dosimetric, and position factors for radiation-induced acute esophagitis in intensity-modulated (chemo)radiotherapy for locally advanced

non-small-cell lung cancer. *Onco Targets Ther.* 2018;11:6167-6175.

doi:10.2147/OTT.S174561

39. R: The R Project for Statistical Computing. Accessed November 26, 2021. <https://www.r-project.org/>
40. Scutari M. Learning Bayesian Networks with the bnlearn R Package. Published online 2009. doi:10.18637/JSS.V035.I03
41. Lee AW, Ng WT, Pan JJ, et al. International Guideline on Dose Prioritization and Acceptance Criteria in Radiation Therapy Planning for Nasopharyngeal Carcinoma. *Int J Radiat Oncol Biol Phys.* 2019;105(3):567-580. doi:10.1016/j.ijrobp.2019.06.2540
42. Mell L. NRG HN004 Randomized Phase II/III Trial of Radiotherapy with Concurrent MEDI4736 (Durvalumab) vs. Radiotherapy with Concurrent Cetuximab in Patients with Locoregionally Advanced Head and Neck Cancer with a Contraindication to Cisplatin. Randomized Phase II/III Trial of Radiotherapy with Concurrent MEDI4736 (Durvalumab) vs. Radiotherapy with Concurrent Cetuximab in Patients with Locoregionally Advanced Head and Neck Cancer with a Contraindication to Cisplatin. Accessed November 26, 2021. <https://www.nrgoncology.org/Clinical-Trials/Protocol/nrg-hn004?filter=nrg-hn004>
43. Lee N. NRG HN001 Randomized Phase II and Phase III Studies of Individualized Treatment for Nasopharyngeal Carcinoma Based on Biomarker Epstein Barr Virus (EBV) Deoxyribonucleic Acid (DNA). Accessed November 26, 2021. <https://www.nrgoncology.org/Clinical-Trials/Protocol/nrg-hn001?filter=nrg-hn001>
44. Trotti A. RTOG 1016 Phase III Trial of Radiotherapy Plus Cetuximab Versus

Chemoradiotherapy in HPV-Associated Oropharynx Cancer. Accessed November 26, 2021.

<https://www.nrgoncology.org/Clinical-Trials/Protocol/rtog-1016?filter=rtog-1016>

45. Rodriguez C. RTOG 1008 A Randomized Phase II/Phase III Study of Adjuvant Concurrent Radiation and Chemotherapy versus Radiation Alone in Resected High-Risk Malignant Salivary Gland Tumors. Accessed November 26, 2021.

<https://www.nrgoncology.org/Clinical-Trials/Protocol/rtog-1008?filter=rtog-1008>

46. Gillison M. RTOG 3504 Safety Evaluations of Nivolumab (Anti-PD-1) Added to Chemoradiotherapy in Patients with Intermediate and High-Risk Local-Regionally Advanced Head and Neck Squamous Cell Carcinoma. Accessed November 26, 2021.

<https://www.rtog.org/Clinical-Trials/Foundation-Studies/3504>

47. Mehta M. NRG BN001 Randomized Phase II Trial of Hypofractionated Dose-Escalated Photon IMRT or Proton Beam Therapy Versus Conventional Photon Irradiation With Concomitant and Adjuvant Temozolomide in Patients With Newly Diagnosed Glioblastoma. Accessed November 26, 2021. <https://www.nrgoncology.org/Clinical-Trials/Protocol/nrg-bn001?filter=nrg-bn001>

48. Grosshans D. NRG BN005 A Phase II Randomized Trial of Proton vs. Photon Therapy (IMRT) for Cognitive Preservation in Patients with IDH Mutant, Low to Intermediate Grade Gliomas. Accessed November 26, 2021. <https://www.nrgoncology.org/Clinical-Trials/Protocol/nrg-bn005?filter=nrg-bn005>

49. Machtay M. RTOG 0920 A Phase III Study of Postoperative Radiation Therapy (IMRT) +/- Cetuximab for Locally-Advanced Resected Head and Neck Cancer. Accessed November

- 26, 2021. <https://www.nrgoncology.org/Clinical-Trials/Protocol/rtog-0920?filter=rtog-0920>
50. Bauman JE, Harris J, Uppaluri R, et al. NRG-HN003: Phase I and Expansion Cohort Study of Adjuvant Pembrolizumab, Cisplatin and Radiation Therapy in Pathologically High-Risk Head and Neck Cancer. *Cancers (Basel)*. 2021;13(12):2882. doi:10.3390/cancers13122882
51. Viswanathan AN, Moughan J, Miller BE, et al. NRG Oncology/RTOG 0921: A phase 2 study of postoperative intensity-modulated radiotherapy with concurrent cisplatin and bevacizumab followed by carboplatin and paclitaxel for patients with endometrial cancer. *Cancer*. 2015;121(13):2156-2163. doi:10.1002/cncr.29337
52. Rogers L. NRG BN003 Phase III Trial of Observation Versus Irradiation for a Gross Totally Resected Grade II Meningioma. Accessed November 26, 2021. <https://www.nrgoncology.org/Clinical-Trials/Protocol/nrg-bn003?filter=nrg-bn003>
53. Ng WT, Lee MCH, Hung WM, et al. Clinical outcomes and patterns of failure after intensity-modulated radiotherapy for nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys*. 2011;79(2):420-428. doi:10.1016/j.ijrobp.2009.11.024
54. Tagliaferri L, Budrukkar A, Lenkowicz J, et al. ENT COBRA ONTOLOGY: the covariates classification system proposed by the Head & Neck and Skin GEC-ESTRO Working Group for interdisciplinary standardized data collection in head and neck patient cohorts treated with interventional radiotherapy (brachytherapy). *J Contemp Brachytherapy*. 2018;10(3):260-266. doi:10.5114/jcb.2018.76982
55. Lancellotta V, Guinot JL, Fionda B, et al. SKIN-COBRA (Consortium for Brachytherapy data Analysis) ontology: The first step towards interdisciplinary standardized data

collection for personalized oncology in skin cancer. *J Contemp Brachytherapy*.

2020;12(2):105-110. doi:10.5114/jcb.2020.94579

56. Schuler T, Kipritidis J, Eade T, et al. Big Data Readiness in Radiation Oncology: An Efficient Approach for Relabeling Radiation Therapy Structures With Their TG-263 Standard Name in Real-World Data Sets. *Adv Radiat Oncol*. 2019;4(1):191-200. doi:10.1016/j.adro.2018.09.013
57. Cardan RA, Covington EL, Popple RA. Technical Note: An open source solution for improving TG-263 compliance. *J Appl Clin Med Phys*. 2019;20(9):163-165. doi:10.1002/acm2.12701
58. Brouwer CL, Steenbakkers RJHM, Bourhis J, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother Oncol*. 2015;117(1):83-90. doi:10.1016/j.radonc.2015.07.041
59. Verbakel WFAR, Doornaert PAH, Raaijmakers CPI, et al. Targeted Intervention to Improve the Quality of Head and Neck Radiation Therapy Treatment Planning in the Netherlands: Short and Long-Term Impact. *Int J Radiat Oncol Biol Phys*. 2019;105(3):514-524. doi:10.1016/j.ijrobp.2019.07.00

Tables/Figures Legends

Legends

Figure 1) Inter-institutional variability in a) naming and b) use of the 13 structures used by 3/5 institutions for at least 50% of plans. Musc_Constricts includes (Musc_Constrict_S, Musc_Constrict_I, Pharynx and PTV-Pharynx). Each structure bar represents the range of values (number of name variants or plans with structure segmented) of the 5 institutions. GlnD_Submands and Parotids indicates contouring both left and right structures.

Figure 2) Variability in contoured structure volumes is evident in a) the wide range of values relative to median. Median volume values are provided adjacent to structure names on the Y axis. b) Three groupings were identified, for low (green shade), moderate (gray shade) and high (no shading) variability of contoured volumes based on K means clustering of volumes according to intra-institutional variability ($\langle(Q3-Q1)/Median\rangle$) and inter-institutional variability ($\langle CHD\rangle$).

Figure 3) Statistical DVH curves illustrating variation of doses for a) GlnD Submand_Low: a structure with low inter-institutional volume variability ($\langle CHD\rangle = 0.25\pm 0.059$) and b) Larynx: a structure with moderate inter-institutional volume variability ($\langle CHD\rangle 0.41\pm 0.1$). Curves show median (dashed line), and ranges encompassing 25%-75% (dark pink), 15%-85% (medium pink), 5% -95% (light pink) of DVH curves from each institution.

Figure 4) Histograms illustrating intra and inter-institutional variation of one DVH metric:

Mean[Gy] for a) GInd_Submand_Low and b) Larynx.

Figure 5) Bayesian network analysis of predictive relationships (strong: solid lines, moderate: dashed lines) among volume based DVH metrics for all institutions.

Journal Pre-proof

Table 1) Comparison of distributions of real-world DVH Metric values to institutional and literature constraints. Substantial variability was noted among institutions for which constraints, and prioritizations were used as part of routine practice. Quantile analysis of (Median [Q1, Q3]) of average values of specific DVH metrics for individual institutions from their “real-world”, routine clinical practice, showed substantial variation in comparison to literature guideline values cited clinical trials and other publications. For some structures, such as Esophagus, and Larynx, “real-world” average values were substantially lower than literature guideline values and routine practice experience suggesting lower constraint values might be warranted.

| OAR Structure | DVH Metric | Planning Constraints (Institution, Priority) | Median [Q1,Q3] | Compare to Literature Guideline Values (with p values for guideline value different from treated values) |
|----------------|-------------|---|----------------|--|
| SpinalCord | D0.03cc[Gy] | < 45Gy (D,1) Max[Gy] < 50Gy (C,1) < 45Gy (B),(E,1) D0.1cc[Gy] < 45Gy (A,1),(B) | 39Gy [36,41] | < 45Gy NRG:1008, 0912, Lee [45,51] (0.006[< 0.001,0.02]) < 48Gy ,NRG:0920, HN003 [49,50] (0.001[<0.001,0.002]) < 50Gy NRG:HN004,1016,1008,3504, BN001, BN005 [42,44-48] (<0.001[<0.001, <0.001]) |
| SpinalCord_PRV | D0.03cc[Gy] | < 50Gy (C,1), (D,1) Max[Gy] < 50Gy (B) < 45Gy (E) D0.1cc[Gy] < | 46Gy [42,48] | < 45 Gy: Lee [41] (0.42 [0.06, 0.06]) |

| | | | | |
|---------------|-------------|--|---------------|---|
| | | 52Gy (C,1) < 50Gy (A,1) | | |
| Brainstem | D0.03cc[Gy] | U 8 (B), (C,1) D0.1cc[Gy] 54 (A,1), (B) | 37Gy [28,42] | < 50Gy ,NRG:HN003 [50] (0.005 [< 0.001, < 0.02]) < 54Gy ,NRG:HN004, BN003, Lee [42,52,41] (0.002 [< 0.001, < 0.006]) < 55Gy ,NRG: BN001, BN005 [47,48] (0.001 [< 0.001, < 0.005]) |
| | V30Gy[%] | < 30% (E,3) | 7.7%[0.6,18] | |
| Brainstem_PRV | D0.03cc[Gy] | U 8 60Gy (B)) 8 54Gy (A,1), (B) | 52Gy[46,56] | <54Gy: Lee[41] (0.43[0.02, 0.43]) |
| Parotid_High | Mean[Gy] | < 26Gy (B,) (C,3), (D,3), (E,3) < 24Gy (A,3) | 30Gy [25,40] | < 26Gy both parotids, Lee [41] (0.009 [0.14,0.02]) |
| Parotid_Low | Mean[Gy] | < 26Gy (B), (C,3), (D,3), (E,3) < 24 (A,3) | 23Gy [16,25] | < 20Gy; <20% long term loss of function, QUANTEC-Deasy [15] (0.03 [0.20,0.004]) < 26Gy both parotids , Lee [41] (0.03 [0.02,0.09]) |
| | V15Gy[%] | < 50% (E,3) | 56% [35,68] | |
| | V15Gy[cc] | | 15cc [7.8,21] | |
| Bone_Mandible | D0.03cc[Gy] | Max[Gy] < 70Gy (E,3) < 66Gy (C,3) D0.1cc[Gy] < 70Gy (A,3) | 70Gy [64,73] | < 66Gy NRG:HN003[50] (0.01[0.09,<0.001]) < 70Gy ,NRG:HN004,3504[42,46] (1.0 [0.003,0.002]) |

| | | | | |
|-----------|-----------|--|-----------------|---|
| | V40Gy[%] | < 40Gy (E,3) | 42% [26,61] | |
| Esophagus | Mean[Gy] | < 45Gy (B) < 30Gy (C,3), (E,3) < 20Gy (A,1) | 21Gy [15,28] | < 30Gy ,NRG:HN004,3504 [42,46] (0.02 [0.008, 0.61]) < 34Gy ; 5-20% acute grade >= 3 Esophagitis, QUANTEC- Werner-Waskik[18] (0.006 [0.003,0.17]) < 35Gy, NRG : HN003 [50] (0.004 [0.002, 0.12]) < 45Gy, larynx cancer NRG : HN003 [50] (0.006 [< 0.001, 0.009]) |
| | V35Gy[%] | | 24% [9.4,41] | < 50% ; > 30% acute grade 2 Esophagitis, QUANTEC- Werner-Waskik[18] (0.21 [<0.001, 0.41]) |
| | V35Gy[cc] | | 3.1cc [1.3,5.3] | |
| Larynx | V50Gy[%] | < 50% (C,3) | 14% [5.3,33] | Median[Gy] < 50Gy risks aspiration , Feng [30] (0.016 [0.002, 0.14]) Median[Gy] < 55Gy risks dysphagia, Akagunduz [28] (0.007 [0.001, 0.05]) |
| | V50Gy[cc] | | 4.5Gy [1.7,11] | |
| | Mean[Gy] | < 45Gy (B) < 43.5Gy (C,3) < 30Gy (E,3) < 20Gy (A,1) | 33Gy [27,41] | < 20Gy , NRG:HN004, 3504 [42,46] (0.04 [0.1, 0.01]) < 35Gy ;glottic, NRG:HN003, Lee [50,41] |

| | | | | |
|-------------------|----------|--|--|---|
| | | | | (0.66 [0.07, 0.29]) < 50Gy risks 30% Aspiration , Mortensen[23] (0.016 [0.002, 0.14]) < 60Gy , NRG:0912 [49] (0.003 [<0.001, 0.02]) |
| Cavity_Oral | Mean[Gy] | < 30Gy (A,3),(B), (C,3)(E,3) | 31Gy [24,40] | < 30Gy, NRG: HN003, HN004,3504 [50,42,46] (0.46 [0.01, 0.002]) < 35Gy, NRG:0912 [49] (0.03 [0.001, 0.03]) <40Gy, Lee[41] (0.03 [0.003,0.5]) |
| | V30Gy[%] | | 48% [27,71] | % [21] (0.002[<0.001,0.95]) |
| | V50Gy[%] | | 11% [1.7,28] | [21] (0.18 [<0.001,0.03]) |
| GInd_Submand_High | Mean[Gy] | < 40Gy (C,3) < 39Gy (E,3) < 30Gy (A,3) | 66Gy [56,69] | |
| GInd_Submand_Low | Mean[Gy] | < 30Gy (A,3), (D,3) < 26Gy (E,3) | 48Gy [34,59] | < 35Gy , Lee[41] (0.07 [0.79, 0.002]) < 39Gy stimulated salivary flow rates recover, Murdoch-Kinch [34] (0.17 [0.23, 0.005]) |
| Eye_(R or L) | Mean[Gy] | | 3Gy [1.5,5.6] R 2.9Gy [1.5,5.2] L | < 35Gy, Lee [41] (<0.001[<0.001,<0.001]) |

| | | | | |
|-------------------|-------------|---|--------------------------------------|--|
| Brain | D1cc[Gy] | < 54Gy (E,3) | 46Gy [37,55] | |
| OpticNrv_(R or L) | D0.03cc[Gy] | < 54Gy (D,1) Max[Gy] 45Gy (B) D0.1cc[Gy] < 54Gy (A,1) | 7.6cc [3.7,14] R 7.3cc [4.1,13] L | Gy Lee [41] (< 0.001[<0.001,<0.001]) |
| OpticChiasm | D0.03cc[Gy] | D0.1cc[Gy] < 54Gy (A,1) | 10cc [4.4,20] | < 54Gy ,NRG:HN004, BN003, Lee [42,52,41] (< 0.001[<0.001,<0.001]) < 55Gy ,NRG:BN01,BN005 [47,48] (< 0.001[<0.001,<0.001]) |
| Cochlea_(R or L) | Mean[Gy] | < 30Gy (D,3) | 9.8Gy [4.7,18] R 9.4Gy [4.5,19] L | <45Gy 30% Sensory neural hearing loss QUANTEC-Bhandare, Lee [20,41] (< 0.001[<0.001,<0.001]) |
| Musc_Constrict_S | Mean[Gy] | < 50Gy (A,3) | 53Gy [45,57] | < 60Gy < 30% aspiration, Mortensen [23] (0.04 [0.1,0.46]) |
| Musc_Constrict_I | Mean[Gy] | < 20Gy (A,1) | 36Gy [29,45] | |
| Pharynx | Mean[Gy] | < 45Gy (B),(C,3) | 48Gy [43,53] | < 45Gy, NRG:HN003,HN004,3504, Lee [50,42,46,41] (0.51 [0.61,0.21]) < 50Gy > 20% Rate Dysphagia and aspiration, QUANTEC-Rancait [16] (0.65[0.13,0.61]) < 60Gy aspirations, Feng [30] (0.04 [0.01,0.26]) |

Figure 1

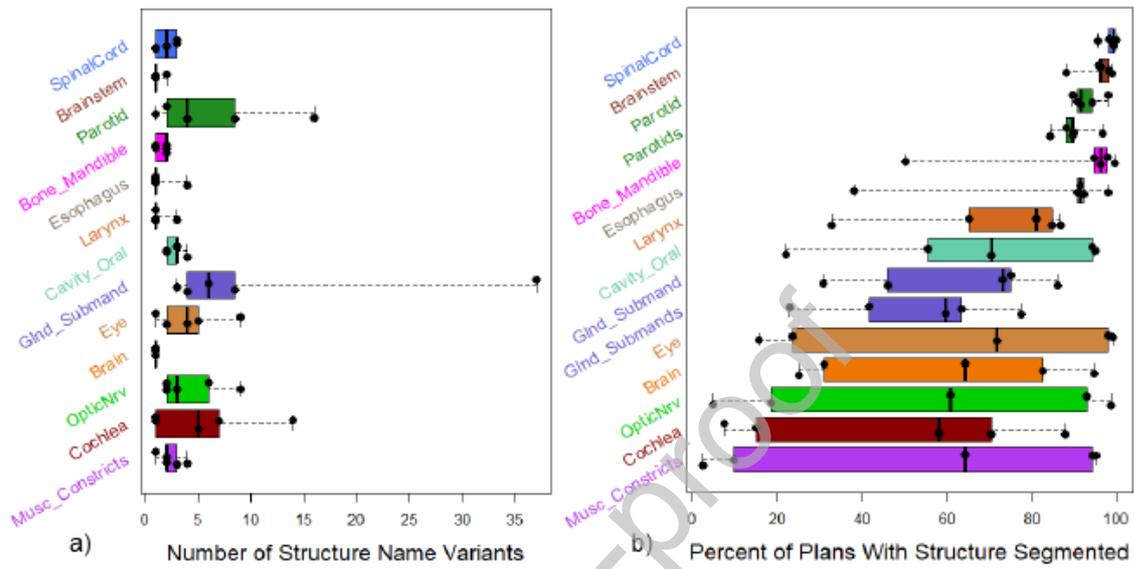


Figure2

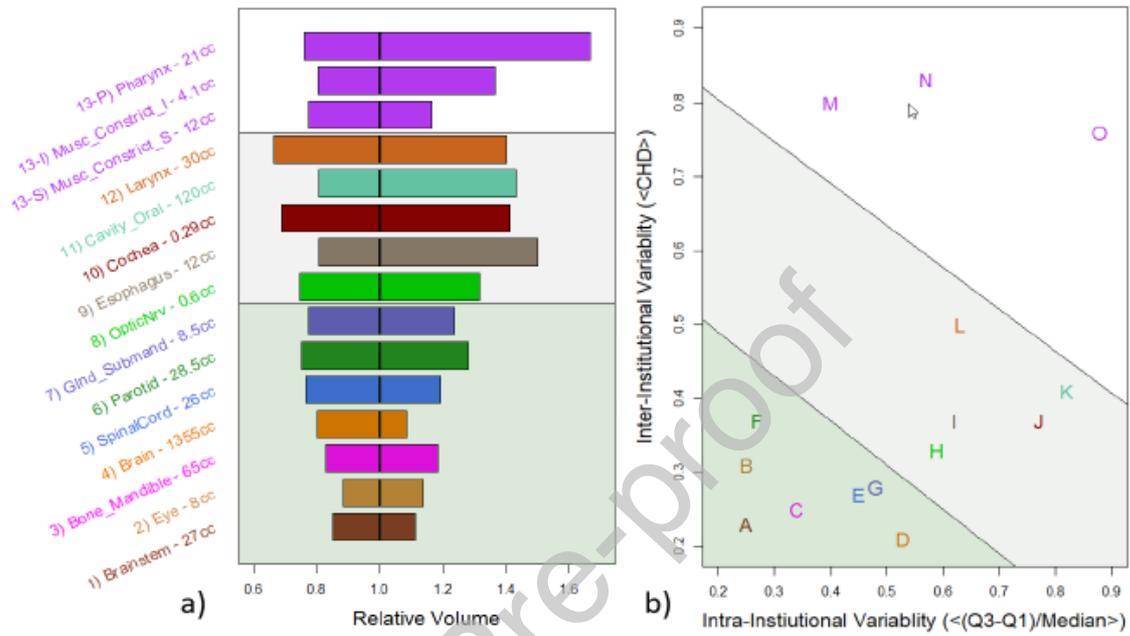


Figure 3

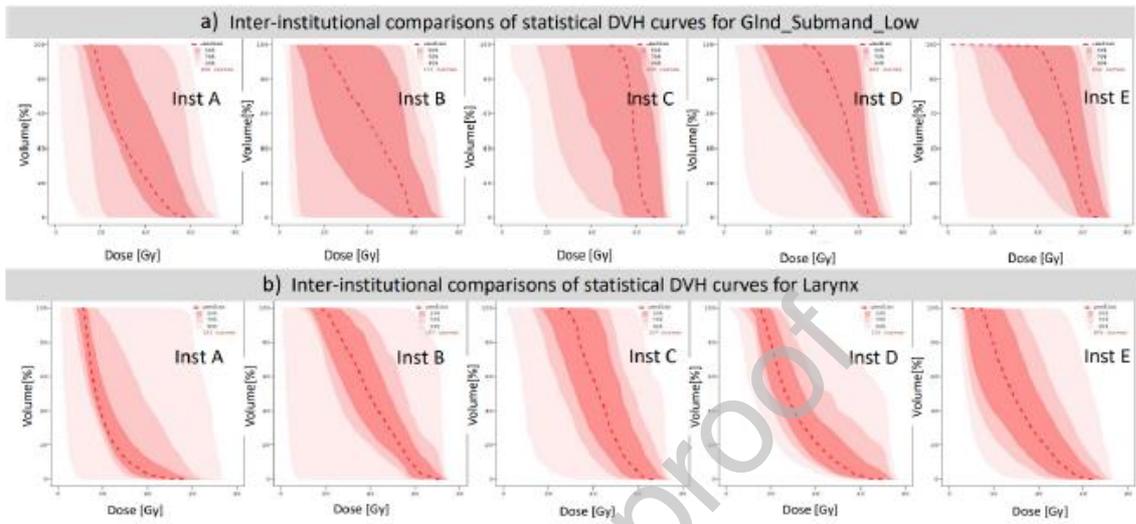


Figure 4

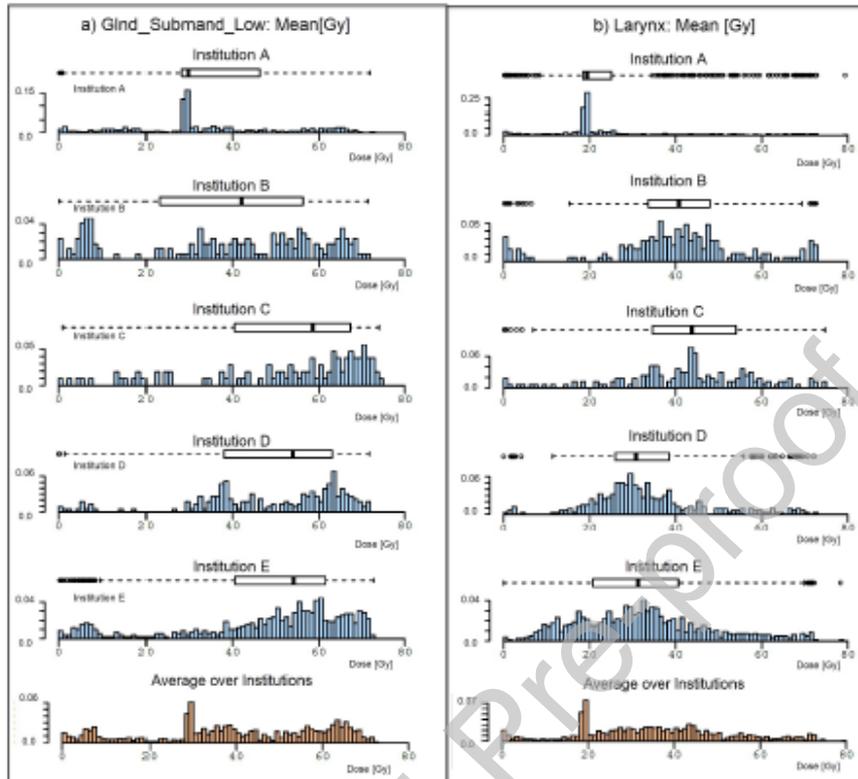


Figure 5

