Scientific Article

Head and Neck Radiation Therapy 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,[a] Michelle Mierzwa, MD,[b] Clifton David Fuller, MD, PhD,[c] Murali Rajaraman, MD,[a] Alex Lin, MD,[d] Andrew MacDonald, MD,[e] Richard Popple, PhD,[e] Ying Xiao, PhD,[d] Lisanne VanDijk, MD,[c] Peter Balter, PhD,[c] Helen Fong,[a] Heping Xu, PhD,[a] Matthew Kovoor,[b] Joonsang Lee, PhD,[b] Arvind Rao, PhD,[b] Mary Martel, PhD,[c] Reid Thompson, MD,[f] Brandon Merz, PhD,[f] John Yao, PhD,[b] and Charles Mayo, PhD[b],*

[a]*Dalhousie University, Halifax, Nova Scotia, Canada;* [b]*University of Michigan, Ann Arbor, Michigan;* [c]*MD Anderson Cancer Center, Houston, Texas;* [d]*University of Pennsylvania, Philadelphia, Pennsylvania;* [e]*University of Alabama, Birmingham, Alabama;* and [f]*University of Oregon Health Sciences Center, Portland, Oregon*

Abstract

Purpose: Outside of randomized clinical trials, it is difficult to develop clinically relevant evidence-based recommendations for radiation therapy (RT) practice guidelines owing to lack of comprehensive real-world data. To address this knowledge gap, we formed the Learning from Analysis of Multicenter Big Data Aggregation 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 (OARs) 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.
Methods and Materials: Dosimetric details of patients with head and neck cancer treated with radiation therapy 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 with 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. Before implementation of TG-263 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.

Conclusions: 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.

Introduction

We know a great deal about the small percentage of patients who are treated on randomized controlled trails but comparatively little about the treatment approaches and outcomes for the large percentage of patients treated in routine practice. Although 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 a real-world setting. There is a need for comprehensive real-world data characterizing efficacy and toxic effects of anticancer treatments such as radiation therapy (RT).[1] Large-scale, real-world data have potential not only to augment clinical trial design and validation[2] but also to improve RT plan quality and patient outcomes by characterizing and reducing practice variability.[3] However, assessment of RT clinical practice patterns is challenged by the complexity of nonstandardized electronic dosimetric data. International efforts are currently underway to promote standardization of RT data, including consensus papers from the American Society for Radiation Oncology (ASTRO) recommending standardized RT normal-tissue contouring, minimum data sets, and synoptic treatment summaries; the TG-263 guideline for standardized RT nomenclature from the American Association of Physicists in Medicine (AAPM); and the ongoing development of Patient Reported Outcome and Big Data guidance documents from the Canadian Partnership for Quality Radiotherapy. These RT society−led quality improvement initiatives are helping to pave the way toward facilitated capture and use of RT "big data."[4-8]

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 toxic effects from treatment.[9-36] As the first of its kind to assess head and neck (H&N) RT "big data," we have formed the Learning from Analysis of Multicentre 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 evidence-based RT plan quality recommendations for structures and DVH metrics for LAMBDA members. The aim of these recommendations 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.

Methods and Materials

Six institutions are currently participants in the LAMBDA consortium: Dalhousie University, University of Michigan, MD Anderson Cancer Centre, University of Pennsylvania, University of Alabama, University of Oregon Health Sciences Center. 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 data sets were anonymized by the submitting institutions before submission to the central institution (University of Michigan) 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 SQL, version 12.0) were used. All analyses were carried out using R, version 3.4.1, a computational statistics software package.[37] Significance of differences in comparison of any values was determined with the Student t test using a threshold of $P = .05$.

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 (eg, Parotid_L and Parotid_R), they were additionally subcategorized according to the relative mean dose (eg, Parotid_High and Parotid_Low).

Statistical DVH curves were used to visualize quantified comparisons of DVH curves among institutions. This provided a graphical means of identifying interinstitutional 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 half of the difference in each histogram bin to calculate the cumulative histogram difference (CHD). For example, if the distributions of structure volume values for 2 institutions were identical, then CHD = 0. If there was no overlap in the distribution of volumes for the 2 institutions, then CHD = 1. The CHD values for each unique pair of institutions were averaged (<CHD>) to quantify the interinstitutional 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 interinstitutional variability (ie, <CHD>) could be calculated. The <CHD> values were used to group structure volumes and DVH metrics according to low (<0.4), moderate (0.4-0.8), and high (>0.8) interinstitutional variability.

Distributions were further summarized by calculating median, first quantile (25% of values ≤Q1), and third quantile (75% of values ≤Q3) values. Intra-institutional variability was quantified by averaging values for (Q3 − Q1)/median (ie, <[Q3 − Q1]/median>). A k-means clustering algorithm was used as an objective means to group structures by <(Q3 − Q1)/median> and <CHD> values, providing a visualization of structure volumes according to the amount of intra- and interinstitution variability.

The RT-planning DVH metrics used by institutions were summarized and compared with the wide range of metrics seen in the literature.[14-15,17,19-20,22,27,29,33,38-50] For structures with high dose constraints, the D0.03cc[Gy] was calculated and compared across institutions. "Real-world" treated values were represented as medians (Q1, Q3), and interinstitutional variability in these dose values was assessed using <CHD>. When only a portion of the structure volume is drawn (eg, entire esophagus vs a portion of esophagus proximal to the target), DVH metrics based on a percentage of the structure (eg, V50Gy[%] the percentage of the structure volume receiving 50 Gy or more) may be less consistent than metrics based on the absolute volume (eg, V50Gy[cc] using TG-263 nomenclature, the volume of the structure in cubic centimeters receiving 50Gy or more). Absolute volume (VxGy[cc]) CHD were compared with percentage volume versions (VxGy[%]), using <CHD> values to confirm whether the absolute volume DVH metric had less interinstitutional variability. Owing to the proximity of structures, DVH metric values for 1 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, version 4.5).[51]

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 interinstitutional variability.

Results

Data were analyzed from 4664 patients from 5 LAMBDA institutions. The average age of patients was 60.1 ± 11 years. Categorized by planning target volume (PTV) dose, the cohort was made up of 2 ± 2% palliative (≤50 Gy), 41 ± 7% adjuvant (>50 Gy and <70 Gy), and 56 ± 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 of 5) did not have systems in place to automate electronic extraction of staging information during the years of the study. Of the 2 institutions able to report staging information, their distributions were stage I (4%, 6%), II (5%, 9%), III (26%, 18%), and IV (65%, 67%).

Variability of OAR nomenclature and contour inclusion

Figure 1 summarizes OAR structures included in the H&N RT plan data sets. Before implementation of TG-263 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 the left parotid. Institution A had the lowest number of name variants per structure (mean ± standard deviation, 1.5 ± 0.8), followed by institutions C (2.0 ± 1.3), D (3.2 ± 2.4), B (3.4 ± 2.5) and E (6.3 ± 8.4). Bilateral OAR structures showed substantial variability in nomenclature (5.9 ± 4.9): optic nerves, 4.4 ± 2.9; lacrimal gland, 3.5 ± 2.4; parotid gland, 6.3 ± 5.7; and submandibular gland, 11.6 ± 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.
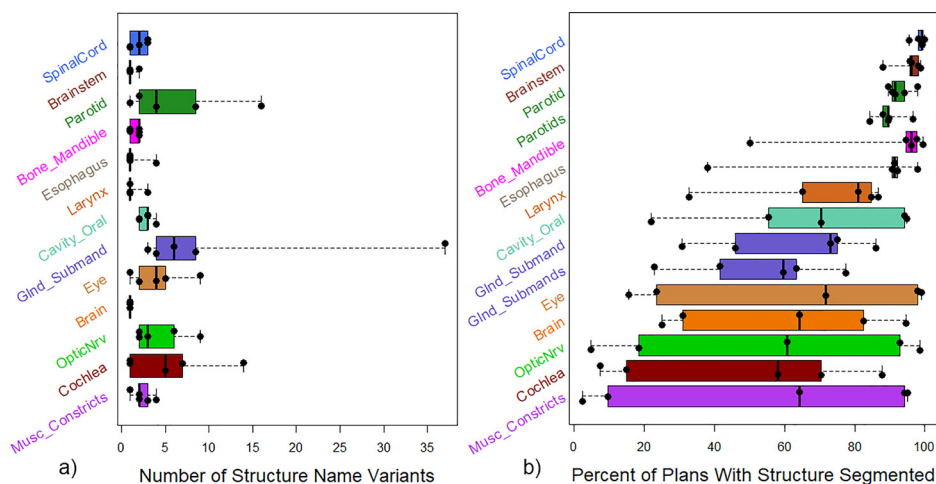
Figure 1 Interinstitutional variability in (a) naming and (b) use of the 13 structures used by 3 of 5 institutions for at least 50% of plans. Musc_Constricts includes Musc_Constrict_S, Musc_Constrict_I, Pharynx, and Pharynx-PTV. Each structure bar represents the range of values (number of name variants or plans with structure segmented) of the 5 institutions. Glnd_Submands and Parotids indicates contouring both left and right structures.

Of the 54 OAR structures used by the institutions, only 2, SpinalCord and at least 1 Parotid, were contoured for ≥90% of all patients treated. Six structures were contoured for ≥90% of patients at 3 of 5 institutions: SpinalCord, Brainstem, at least 1 Parotid, Bone_Mandible, Larynx, and Esophagus. After lowering the threshold to 50% of patients treated at 3 of 5 institutions, an additional 7 structures were identified: oral cavity, brain, and at least 1 of the following bilateral structures: submandibular glands, eyes, cochleas, optic nerves, and a structure to monitor dose to constrictor muscles (eg, Pharynx, Musc_Constrict_S, Musc_Constrict_I, Musc_Constric-PTV). For planning OAR volumes (PRVs), only the SpinalCord_PRV was routinely (≥75% of patients) contoured in the majority of institutions (≥3 of 5). Both parotids were contoured in at least 84.3% of plans (institution A), whereas 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 1 institution (E), whereas 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% of cases) contoured both superior and inferior constrictor muscles, D contoured Musc_Constric-PTV in 95% of cases, and C and B contoured the pharynx in 9% and 64% of cases, respectively. Institution E contoured constrictors in only 2.4% of cases. Less than 20% of total RT plans included contouring of both parotid glands (Parotids), both submandibular glands (Glnd_Submands), and the muscle constrictors (Musc_Constrict).

Variability of OAR volumes

There was substantial variability among institutions not only 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%, and 58% at A, C, D, and E, respectively). On the other hand, institution B contoured the cochlea in only 15.4% of cases but contoured the middle ear and inner ear separately for 65% of RT plans. Contouring of the brain (82.6%, 64.3%, 94.6%, 25%, and 30.9% at A-E, respectively) or temporal lobe (15%, 0%, 2.4%, 79.0%, and 0.2% at A-E, respectively) varied substantially. One institution routinely contoured only the subvolume of OARs outside of the PTV (eg, 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 toxic effects.

Figure 2 summarizes the analysis of institutional norms for OAR structure volumes. Among the 13 OAR volumes segmented on ≥50% of patients for the majority of institutions, 6 (Brainstem, Eye, Bone_Mandible, Parotid, submandibular glands [Glnd_Submands], and SpinalCord) had low interinstitution variability (<CHD> <0.4; P < .05). Volumes for OpticNrv, Esophagus, Cochlea, Cavity_Oral, and Larynx had moderate interinstitution variability. Volumes for constrictor muscles had the highest interinstitution variability (>0.8 <CHD>).

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 ≥3 of the
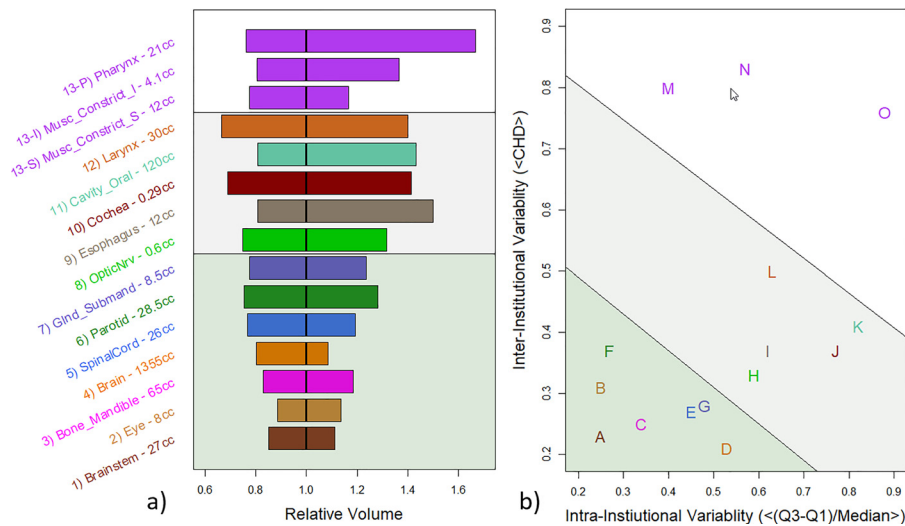
Figure 2 Variability in contoured structure volumes is evident in **(a)** the wide range of values relative to median. Median volume values are provided adjacent to structure names on the Y axis. **(b)** Three groupings were identified for low (green shading), moderate (gray shading), and high (no shading) variability of contoured volumes based on k-means clustering of volumes according to intrainstitutional variability ($<[Q3 − Q1]/\text{median}>$) and interinstitutional variability (<CHD>).

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 the 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). Among the institutions that prioritized high dose constraints for the critical structures of SpinalCord, Brainstem, and OpticNrvs and OpticChiasm, priority 1 was assigned for these structures. There was interinstitutional variability in inclusion of priority levels for DVH metrics of other structures and critical structures of SpinalCord_PRV (3 of 4 institutions), OpticNrv/OpticChiasm (2 of 4 institutions), and Brainstem PRV (1 of 4 institutions). There were unique patterns of practice identified for individual institutions, such as institution A, which assigned priority 1 to Larynx and Musc_Contrict_I, whereas other institutions either did not assign priority level to the planning constraints of those structures or assigned them a level of 3.

Real-world treated values were substantially less than institutional or literature-based constraints for some structures.(10=58) The Q3 values were significantly lower than priority 1 constraints for SpinalCord: D0.03cc[Gy] ≤ 45, Brainstem: D0.03cc[Gy] ≤ 54, and OpticNrv/OpticChiasm: D0.03cc[Gy] < 54. For priority 3 constraints, Q3 values were significantly lower: Larynx:Mean[Gy] ≤ 45, Cochlea:Mean[Gy] ≤ 30, and Esophagus:Mean[Gy] ≤ 45. Median values for priority 3 constraints were regularly exceeded for certain OARs (eg, 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 interinstitutional variability (<0.4 <CHD>) compared with percentage volume versions (VxGy[%]) that showed moderate interinstitution 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 the effect of contour practice variability for these structures.

Contour variability is but 1 factor contributing to RT plan variability, as evidenced by statistical DVH curves (Fig. 3) showing dose variations not only for a structure of moderate volume variability (larynx) but also for Glnd_Submand_Low, a structure with low interinstitutional volume variability. Histogram analysis provides visualization of the intra- and interinstitutional variability for a single DVH metric (mean) for structures such as Glnd_Submand_Low or Larynx (Fig. 4). Assessment of only an average median value or quantiles (<median [Q1, Q3]>) of 1 metric such as mean (Table 1: Glnd_Submand_Low:Mean[Gy]< 48 (34, 59) and Larynx:<Mean [Gy]< 33 (27, 41) may lead to underappreciation 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 the 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 (Fig. 5) identified strong predictive associations among DVH metrics of various

Table 1   Comparison of distributions of real-world DVH metric values to institutional and literature constraints*

OAR structure	DVH metric	Planning constraints (institution, priority)	"Real-world" treated values, median (Q1, Q3)	Comparison with literature guideline values (with P values for the guideline value different from "real-world" treated values)
SpinalCord	D0.03cc[Gy]	< 45 Gy (D,1) Max[Gy] < 50 Gy (C, 1) < 45 Gy (B), (E, 1) D0.01cc[Gy] < 45 Gy (A, 1), (B)	39 Gy (36, 41)	< 45 Gy, NRG:1008, 0921, Lee[39,43,49] (0.006 [< 0.001, 0.02]) < 48 Gy, NRG:0920, HN003[47,49] (0.001 [< 0.001, 0.002]) < 50 Gy, NRG:HN004,1016,1008,3504, BN001, BN005[40,42-46] (<0.001 [< 0.001, <0.001])
SpinalCord_PRV	D0.03cc[Gy]	< 50 Gy (C, 1), (D, 1) Max[Gy] < 50 Gy (B) < 45 Gy (E) D0.01cc[Gy] < 52 Gy (C, 1) < 50 Gy (A, 1)	46 Gy (42, 48)	< 45 Gy, Lee[39] (0.42 [0.06, 0.06])
Brainstem	D0.03cc[Gy]	≤ 54 (D, 1) Max[Gy] ≤ 54 (B), (C, 1) D0.01cc[Gy] ≤ 54 (A, 1), (B)	37 Gy (28, 42)	< 50 Gy, NRG:HN003[48] (0.005 [< 0.001, < 0.02]) < 54 Gy, NRG:HN004, BN003, Lee[39,40,50] (0.002 [< 0.001, < 0.006]) < 55 Gy, NRG: BN001, BN005[46,47] (0.001 [< 0.001, < 0.005])
	V30Gy[%])	< 30% (E, 3)	7.7% (0.6, 18)	
Brainstem_PRV	D0.03cc[Gy]	Max[Gy] ≤ 60 Gy (B) D0.01cc[Gy] ≤ 54 Gy (A, 1), (B)	52 Gy (46, 56)	< 54 Gy, Lee[39] (0.43 [0.02, 0.43])
Parotid_High	Mean[Gy]	< 26 Gy (B), (C, 3), (D, 3), (E, 3) < 24 Gy (A, 3)	30 Gy (25, 40)	< 26 Gy, both parotids, Chen[38] Lee[39] (0.009 [0.14, 0.02])
Parotid_Low	Mean[Gy]	< 26 Gy (B), (C, 3), (D, 3), (E, 3) < 24 Gy (A, 3)	23 Gy (16, 25)	< 20 Gy, <20% long-term loss of function, QUANTEC-Deasy[14] (0.03 [0.20, 0.004]) < 26 Gy, both parotids, Lee[39] (0.03 [0.02, 0.09])
	V15Gy[%]	< 50% (E, 3)	56% (35, 68)	
	V15Gy[cc]		15 cc (7.8, 21)	
Bone_Mandible	D0.03cc[Gy]	Max[Gy] < 70 Gy (E, 3) < 66 Gy (C, 3) D0.01cc[Gy] < 70 Gy (A, 3)	70 Gy (64, 73)	< 66 Gy, NRG:HN003[48] (0.01 [0.09, <0.001]) < 70 Gy, NRG:HN004,3504[40,44] (1.0 [0.003, 0.002])
	V40Gy[%]	< 40% (E, 3)	42% (26, 61)	
Esophagus	Mean[Gy]	< 45 Gy (B) < 30 Gy (C, 3), (E, 3) < 20 Gy (A, 1)	21 Gy (15, 28)	< 30 Gy, NRG:HN004,3504[40,44] (0.02 [0.008, 0.61]) < 34 Gy; 5%-20% acute grade ≥3 esophagitis, QUANTEC-Werner-Waskik[17] (0.006 [0.003, 0.17]) < 35 Gy, NRG: HN003[48] (0.004 [0.002, 0.12]) < 45 Gy, larynx cancer, NRG: HN003[48] (0.006 [<0.001, 0.009])
	V35Gy[%]		24% (9.4, 41)	< 50%; >30% acute grade ≥2 esophagitis, QUANTEC-Werner-Waskik[17] (0.21 [<0.001, 0.41])
	V35Gy[cc]		3.1 cc (1.3, 5.3)	

Table 1 (*Continued*)

OAR structure	DVH metric	Planning constraints (institution, priority)	"Real-world" treated values, median (Q1, Q3)	Comparison with literature guideline values (with *P* values for the guideline value different from "real-world" treated values)
Larynx	V50Gy[%]	< 50% (C, 3)	14% (5.3, 33)	Median (Gy) < 50 Gy risks aspiration, Feng[29] (0.016 [0.002, 0.14]) Median (Gy) < 55 Gy risks dysphagia, Akagunduz[27] (0.007 [0.001, 0.05])
	V50Gy[cc]		4.5 cm³ (1.7, 11)	
	Mean[Gy]	< 45 Gy (B) < 43.5 Gy (C, 3) < 30 Gy (E, 3) < 20 Gy (A, 1)	33 Gy (27, 41)	< 20 Gy, NRG:HN004, 3504[40,44] (0.04 [0.1, 0.01]) < 35 Gy, glottic, NRG:HN003, Lee[39,48] (0.66 [0.07, 0.29]) < 50 Gy risks 30% Aspiration, Mortensen[22] (0.016 [0.002, 0.14]) < 60 Gy, NRG:0912[47] (0.003 [<0.001, 0.02])
Cavity_Oral	Mean[Gy]	< 30 Gy (A, 3), (B), (C, 3), (E, 3)	31 Gy (24, 40)	< 30 Gy, NRG: HN003, HN004,3504[40,44,48] (0.46 [0.01, 0.002]) < 35 Gy, NRG:0912[47] (0.03 [0.001, 0.03]) <40 Gy, Lee[39] (0.03 [0.003, 0.5])
	V30Gy[%]		48% (27, 71)	≤ 71.8% grade ≥3 acute toxicity, Li[20] (0.002 [<0.001, 0.95]
	V50Gy[%]		11% (1.7, 28)	≤ 14.3% grade ≥3 acute toxicity, Li[20] (0.18 [<0.001, 0.03])
Glnd_Submand_High	Mean[Gy]	< 40 Gy (C, 3) < 39 Gy (E, 3) < 30 Gy (A, 3)	66 Gy (56, 69)	
Glnd_Submand_Low	Mean[Gy]	< 30 Gy (A, 3), (D, 3) < 26 Gy (E, 3)	48 Gy (34, 59)	< 35 Gy, Lee[39] (0.07 [0.79, 0.002]) < 39 Gy stimulated salivary flow rates recover, Murdoch-Kinch[33] (0.17 [0.23, 0.005])
Eye_(R or L)	Mean[Gy]		3 Gy (1.5, 5.6) R 2.9 Gy (1.5, 5.2) L	< 35 Gy, Lee[39] (<0.001 [<0.001, <0.001])
Brain	D1cc[Gy]	< 54 Gy (E, 3)	46 Gy (37, 55)	
OpticNrv_(R or L)	D0.03cc[Gy]	< 54 Gy (D, 1) Max[Gy] ≤ 45 Gy (B) D0.1cc[Gy] < 54 Gy (A, 1)	7.6 Gy (3.7, 14) R 7.3 Gy (4.1, 13) L	≤ 54 Gy Lee[39] (< 0.001 [<0.001, <0.001])
OpticChiasm	D0.03cc[Gy]	D0.1cc[Gy] < 54 Gy (A, 1)	10 Gy (4.4, 20)	< 54 Gy, NRG:HN004, BN003, Lee[39,40,50] (< 0.001 [<0.001, <0.001]) < 55 Gy, NRG:BN001, BN005[45,46] (< 0.001 [<0.001, <0.001])
Cochlea_(R or L)	Mean[Gy]	< 30 Gy (D, 3)	9.8 Gy (4.7, 18) R 9.4 Gy (4.5, 19) L	<45 Gy, 30% sensory neural hearing loss QUANTEC-Bhandare, Lee[19,39] (< 0.001 [<0.001, <0.001])

Table 1 (*Continued*)

OAR structure	DVH metric	Planning constraints (institution, priority)	"Real-world" treated values, median (Q1, Q3)	Comparison with literature guideline values (with P values for the guideline value different from "real-world" treated values)
Musc_Constrict_S	Mean[Gy]	< 50 Gy (A, 3)	53 Gy (45, 57)	< 60 Gy <30% aspiration, Mortensen[22] (0.04 [0.1, 0.46])
Musc_Constrict_I	Mean[Gy]	< 20 Gy (A, 1)	36 Gy (29, 45)	
Pharynx	Mean[Gy]	< 45 Gy (B), (C, 3)	48 Gy (43, 53)	< 45 Gy, NRG:HN003, HN004, 3504, Lee[39,40,44,48] (0.51 [0.61, 0.21]) < 50 Gy > 20% rate dysphagia and aspiration, QUANTEC-Rancait[15] (0.65 [0.13, 0.61]) < 60 Gy aspirations, Feng[29] (0.04 [0.01, 0.26])

Abbreviations: DVH = dose volume histogram; OAR = organ at risk; Gy = Gray; L = left; R = right; PRV = planning organ at risk volume; QUANTEC = Quantitative Analysis of Normal Tissue Effects in the Clinic.

* Substantial variability was noted among institutions for which constraints and prioritizations were used as part of routine practice. Quantile analysis of the median (Q1, Q3) values of specific DVH metrics for individual institutions from their "real-world," routine clinical practice showed substantial variation in comparison with literature guideline values cited in clinical trials and other publications. For some structures, such as the esophagus and larynx, "real-world" average values were substantially lower than literature guideline values, and routine practice experience suggested lower constraint values might be warranted.

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 multistructure models of toxic effects.

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.
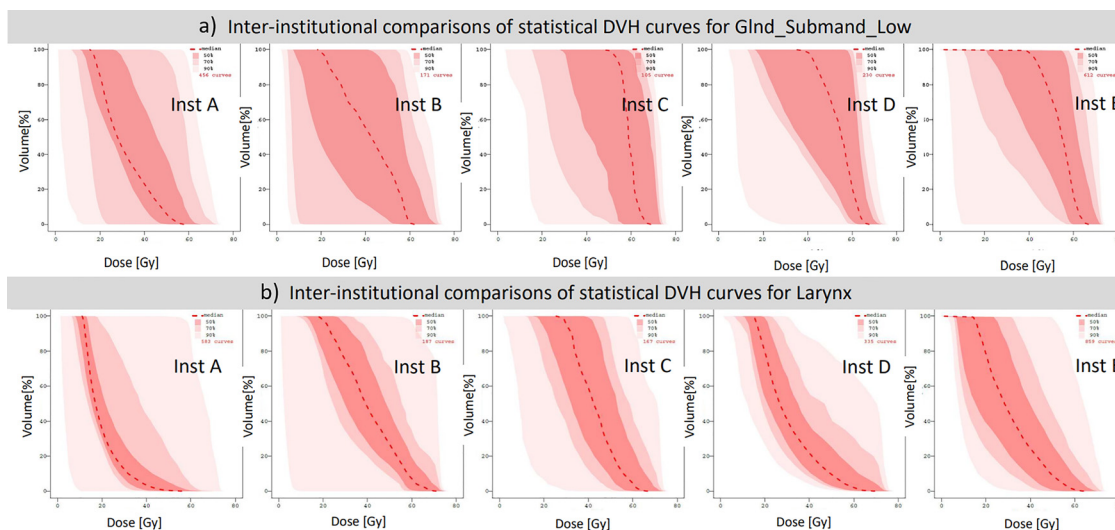

Figure 3 Statistical dose volume histogram (DVH) curves illustrating variation of doses for **(a)** Glnd Submand_Low: a structure with low interinstitutional volume variability (<CHD> = 0.25 ± 0.059) and **(b)** Larynx: a structure with moderate interinstitutional volume variability (<CHD> 0.41 ± 0.1). The median (dashed line) and ranges encompassing 25% to 75% (dark pink), 15% to 85% (medium pink), and 5% to 95% (light pink) of the DVH curves from each institution are shown.
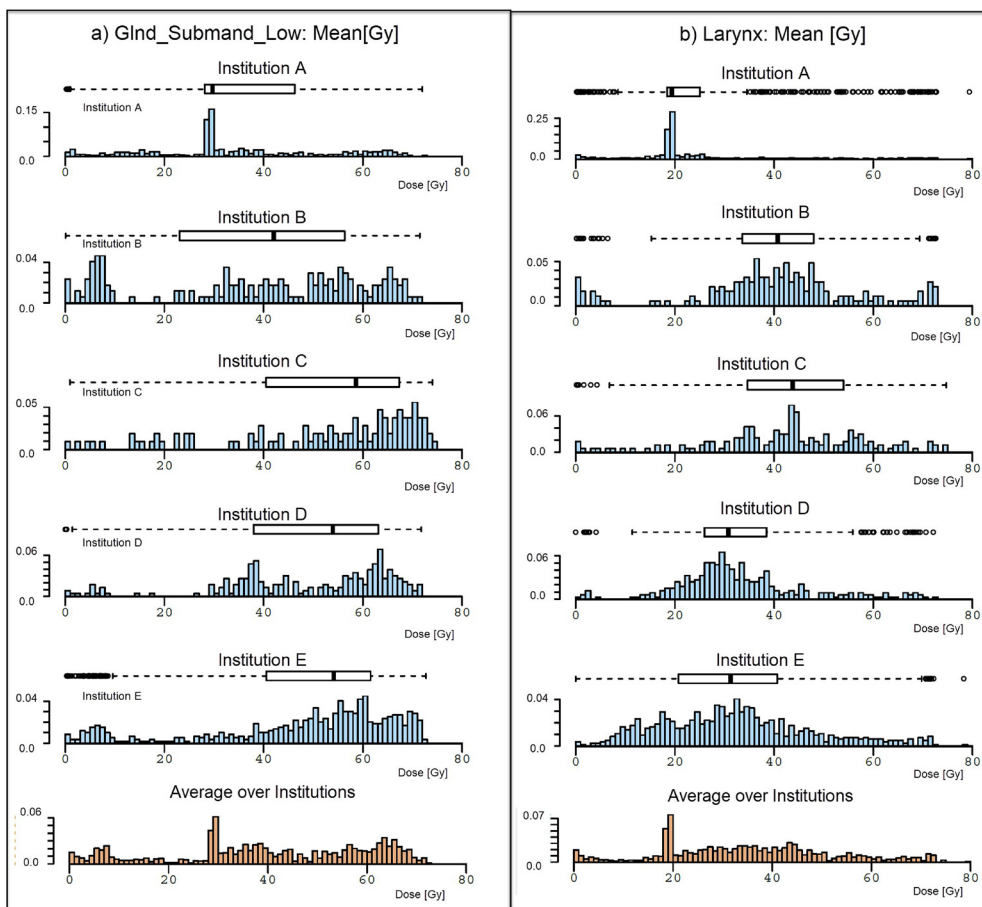
Figure 4 Histograms illustrating intrainstitutional and interinstitutional variation of 1 dose volume histogram metric: the mean (Gy) for **(a)** Glnd_Submand_Low and **(b)** Larynx.

- 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, clinical target volume, and gross tumor volume) names acceptable at each institution.
- As a means of ensuring complete data sets, include the 13 structures contoured on ≥50% of patients in the majority of institutions (≥3 of 5): brain(Brain),
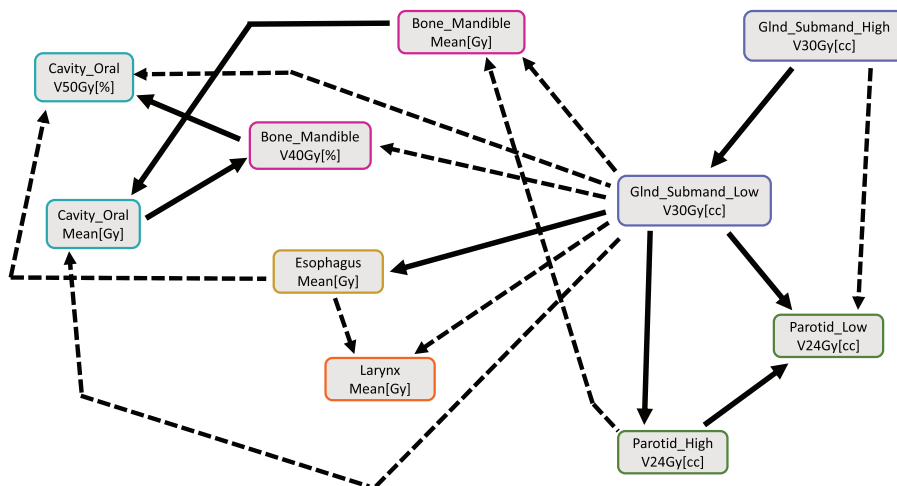

Figure 5 Bayesian network analysis of predictive relationships (strong: solid lines; moderate: dashed lines) among volume-based dose volume histogram metrics for all institutions.

brainstem (Brainstem), spinal cord (SpinalCord), eyes (Eye_L, Eye_R), cochleas (Cochlea_L, Cochela_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 the disease site treated, and all OARs should be contoured following published atlases.[52]

- In data pooling applications, provide at least the minimum set of 18 DVH metrics for reporting that were identified for these 13 structures (Table 1).
- Consistent with guidelines, the 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, and Glnd_Submand_R), include both left and right structures if present (ie, 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.[23,26,31]
- 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.[39]
- Consider reducing constraint values for DVH metrics where median and Q3 values (Table 1) are well below standard limits set in the literature (eg, Esophagus).

Discussion

To our knowledge, the present study was the first to use a large combined data set (>4000 patients) drawn from "real-world" H&N cancer RT practice to quantify interinstitutional variability in routinely segmented OARs 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 Canadian Partnership for Quality Radiotherapy, 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's and ASTRO's combined effort to develop an operational ontology for radiation oncology, which includes professional society—endorsed standardizations like TG-263. As such, this research effort aimed to identify gaps in clinical practice that need to be addressed 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),[53-54] paved the way in H&N ontology work to standardize data collection for patients with H&N cancer 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 a focus on technical approaches to either efficiently relabel retrospective data or improve TG-263 compliance going forward.[55,56] The current study details the significant variability in OAR nomenclature that was present before TG-263 implementation, with certain structures such as submandibular glands having more than 35 name variants across institutions.

Significant interinstitutional 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.[4] In the present study, fewer 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 Glnd_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 Radiation Therapy Oncology Group. As preliminary data, the current study's results allow for inferences to be made with respect to OAR contour patterns of practice based on the volume of OARs determined from DVH metrics. Although it is inferred that practice variability exists for contours of certain structures (eg, the larynx), it is also hypothesized that there are factors beyond contour variation that may affect dose constraints achieved, given that significant interinstitutional

variability was found for DVH metrics of OARs that had similar (eg, submandibular glands) or dissimilar (eg, larynx) volumes across institutions.

These results are consistent with a recent report of the 15 Dutch radiation oncology institutions showing large interinstitutional variations in PTV and OAR dosimetry of a benchmarking test case of 1 H&N RT plan with 6 OAR structures.[57] Although OAR sparing improved through collaborative iterations of contour and plan comparisons, unexplained interinstitutional 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 interinstitutional variability of such OAR prioritization for critical structures such as the spinal cord PRV and brain stem PRV and for structures such as the larynx. Even if OAR constraints are being prioritized, there may be a 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 OARs such as the esophagus that were well below limits set by LAMBDA institutions and published recommendations such as QUANTEC. With the guiding principle of as low as reasonably achievable, 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 Glnd_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 toxic effects, 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.

Although Big Data efforts have numerous advantages, there are limitations that 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 the AAPM's soon-to-be-released oncology ontology will help inform all institutions with regard to standardized data capture of critical data elements beyond dosimetric data that are 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? Although 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 interinstitutional 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 to facilitate future studies of dosimetric OAR data and patient outcomes. Whereas clinical trial results have shown that plan-quality variability negatively affects patient survival,[3] future Big Data studies must investigate whether dosimetric constraint achievement correlates with decreased toxic effects 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;102:924–933.
2. 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.
3. 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;28:2996–3001.
4. 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;9:65–72.

5. 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;10:304−307.

6. 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;9:395−401.

7. 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;10:389−401.

8. 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;100:1057−1066.

9. 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−S160.

10. 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;2:503−514.

11. 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−S19.

12. Mayo C, Yorke E, Merchant TE. Radiation associated brainstem injury. *Int J Radiat Oncol Biol Phys*. 2010;76(3 suppl):S36−S41.

13. 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−S49.

14. 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−S63.

15. 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−S69.

16. 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−S27.

17. 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−S93.

18. 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−S35.

19. Bhandare N, Jackson A, Eisbruch A, et al. Radiation therapy and hearing loss. *Int J Radiat Oncol Biol Phys*. 2010;76(3 suppl):S50−S57.

20. 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.

21. 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;13:194.

22. 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;107:288−294.

23. 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:e93−e99.

24. 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;60:1425−1439.

25. 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;53:23−28.

26. 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;76:403−409.

27. 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;128:73−84.

28. 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;128:442−451.

29. 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;68:1289−1298.

30. 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.

31. 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;98:1022−1027.

32. Nutting CM, Morden JP, Harrington KJ, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): A phase 3 multicentre randomised controlled trial. *Lancet Oncol*. 2011;12:127−136.

33. 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;72:373−382.

34. 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;44:144−149.

35. 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;126:68−74.

36. 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.

37. R: The R Project for Statistical Computing. Available at: https://www.r-project.org/. Accessed November 26, 2021.

38. 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;3:111−120.

39. 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;105:567−580.

40. 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. Available at: https://www.nrgoncology.org/Clinical-Trials/Protocol/nrg-hn004?filter=nrg-hn004. Accessed November 26, 2021.

41. 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). Available at: https://www.nrgoncology.org/Clinical-Trials/Protocol/nrg-hn001?filter=nrg-hn001. Accessed November 26, 2021.

42. Trotti A. RTOG 1016 phase III trial of radiotherapy plus cetuximab versus chemoradiotherapy in HPV-associated oropharynx cancer. Available at: https://www.nrgoncology.org/Clinical-Trials/Protocol/rtog-1016?filter=rtog-1016. Accessed November 26, 2021.

43. 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. Available at: https://www.nrgoncology.org/Clinical-Trials/Protocol/rtog-1008?filter=rtog-1008. Accessed November 26, 2021.

44. 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. Available at: https://www.rtog.org/Clinical-Trials/Foundation-Studies/3504. Accessed November 26, 2021.

45. 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. Available at: https://www.nrgoncology.org/Clinical-Trials/Protocol/nrg-bn001?filter=nrg-bn001. Accessed November 26, 2021.

46. 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. Available at: https://www.nrgoncology.org/Clinical-Trials/Protocol/nrg-bn005?filter=nrg-bn005. Accessed November 26, 2021.

47. Machtay M. RTOG 0920 A phase III study of postoperative radiation therapy (IMRT) +/- cetuximab for locally-advanced resected head and neck cancer. Available at: https://www.nrgoncology.org/Clinical-Trials/Protocol/rtog-0920?filter=rtog-0920. Accessed November 26, 2021.

48. 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;13:2882.

49. 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;121:2156–2163.

50. Rogers L. NRG BN003 phase III trial of observation versus irradiation for a gross totally resected grade II meningioma. Available at: https://www.nrgoncology.org/Clinical-Trials/Protocol/nrg-bn003?filter=nrg-bn003. Accessed November 26, 2021.

51. Scutari M. Learning Bayesian networks with the bnlearn R package. Available at: https://www.jstatsoft.org/article/view/v035i03. Accessed April 19, 2022.

52. 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;117:83–90.

53. 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;10:260–266.

54. 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;12:105–110.

55. 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;4:191–200.

56. Cardan RA, Covington EL, Popple RA. Technical note: An open source solution for improving TG-263 compliance. J Appl Clin Med Phys. 2019;20:163–165.

57. Verbakel WFAR, Doornaert PAH, Raaijmakers CPJ, 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;105:514–524.