Advances in Radiation Oncology (2020) **5**, 1324-1333

advances in radiation oncology

www.advancesradonc.org

**Physics Contribution**

# An Automated Workflow to Improve Efficiency in Radiation Therapy Treatment Planning by Prioritizing Organs at Risk

**Eric Aliotta, PhD,* Hamidreza Nourzadeh, PhD, Wookjin Choi, PhD, Victor Gabriel Leandro Alves, PhD, and Jeffrey V. Siebers, PhD**

*Department of Radiation Oncology, University of Virginia, Charlottesville, Virginia*

Received 22 January 2020; revised 15 April 2020; accepted 16 June 2020

**Abstract**

**Purpose:** Manual delineation (MD) of organs at risk (OAR) is time and labor intensive. Auto-delineation (AD) can reduce the need for MD, but because current algorithms are imperfect, manual review and modification is still typically used. Recognizing that many OARs are sufficiently far from important dose levels that they do not pose a realistic risk, we hypothesize that some OARs can be excluded from MD and manual review with no clinical effect. The purpose of this study was to develop a method that automatically identifies these OARs and enables more efficient workflows that incorporate AD without degrading clinical quality.

**Methods and Materials:** Preliminary dose map estimates were generated for n = 10 patients with head and neck cancers using only prescription and target-volume information. Conservative estimates of clinical OAR objectives were computed using AD structures with spatial expansion buffers to account for potential delineation uncertainties. OARs with estimated dose metrics below clinical tolerances were deemed low priority and excluded from MD and/or manual review. Final plans were then optimized using high-priority MD OARs and low-priority AD OARs and compared with reference plans generated using all MD OARs. Multiple different spatial buffers were used to accommodate different potential delineation uncertainties.

**Results:** Sixty-seven out of 201 total OARs were identified as low-priority using the proposed methodology, which permitted a 33% reduction in structures requiring manual delineation/review. Plans optimized using low-priority AD OARs without review or modification met all planning objectives that were met when all MD OARs were used, indicating clinical equivalence.

**Conclusions:** Prioritizing OARs using estimated dose distributions allowed a substantial reduction in required MD and review without affecting clinically relevant dosimetry.



## Introduction

Organ-at-risk (OAR) delineation is a highly time- and resource-intensive component of the radiation therapy (RT) treatment planning workflow. This is due in large part to the need for manual delineation (MD) of OAR volumes by trained experts. Computer-aided auto-delineation (AD) algorithms are widely available and can save

Sources of support: NIH R01CA222216-01A1.

Disclosures: J.V.S. reports grants and nonfinancial support from Philips Medical Systems and grants from Varian Medical Systems, outside the submitted work.

Research data are not available at this time.

* Corresponding author: Eric Aliotta, PhD; E-mail: eric.aliotta@virginia.edu

https://doi.org/10.1016/j.adro.2020.06.012

time by reducing the need for MD.[1-4] However, despite advancement in the quality of AD algorithms, manual review and modification of AD volumes are still typically performed in the clinical setting to ensure accuracy before planning.[1] This additional step can substantially mitigate the potential efficiency gains of AD and stifle adaptive planning workflows.

Although accurate OAR delineation is an important factor in generating and evaluating RT treatment plans,[5] the level of required accuracy is likely to vary from case to case. This is because dosimetric planning objectives are specific to each OAR and achievable OAR doses depend heavily upon patient-specific factors (eg, prescription dose, tumor size/location, relative OAR positions, beam angles, etc). It is thus not clear when delineation inaccuracies are or are not clinically meaningful. In general, OARs that are sufficiently far from target volumes that they do not approach their organ-specific dose tolerances should not pose a realistic risk, regardless of reasonable delineation inaccuracies. We thus hypothesize that some OARs can be excluded from MD or manual review in many cases with no clinical effect. Automatically identifying these OARs before treatment planning should allow clinics to more fully realize the efficiency gain provided by AD.

The purpose of this work was to develop an automated workflow that identifies OARs that can be excluded from MD or manual review (ie, auto-contours can be used "as-is") without affecting plan quality. Our method combined AD OARs with an automatically generated dose map estimate to make conservative predictions of organ-specific plan quality metrics (PQMs). These predictions were used to identify structures that are at risk of exceeding tolerances in the final treatment plan (ie, high-priority OARs) and those that are not (ie, low-priority OARs). We implemented this method in a cohort of patients with head and neck cancers due to the notable OAR delineation burden in this treatment site and quantified the resulting reduction of delineations requiring manual interaction. The adequacy of priority designations was evaluated in terms of their dosimetric effect on final plan dosimetry compared with reference plans generated using all MD OARs during optimization.

## Methods and Materials

This study was performed using computed tomography (CT) images and clinically used MD OAR structure sets from 10 patients with head and neck cancers who were previously treated at our center. Patient characteristics are described in Table 1. The local institutional review board approved the retrospective use of these data.

The required inputs to the proposed workflow are CT simulation images, planning target volumes (PTVs), dose prescriptions, and dosimetric OAR objectives. The workflow can therefore be performed after the physician has finalized their treatment strategy but before OAR delineation or interactive treatment planning is performed. This method is diagrammed in Figure 1 and consists, broadly, of 3 fully automated steps that are described in detail in the following sections: (1) dose map estimation; (2) OAR auto-delineation; and (3) OAR PQM estimation. Note that in the present work, each of these steps was performed using software that is used clinically at our institution but can easily be replaced with other methods.

### Dose map estimation

An initial dose map estimate ($D_{est}$) was generated for each case by creating a 2 full-arc volumetric modulated arc therapy treatment plan using only target volumes and target-based planning structures for optimization. Three target-based planning structures were used—the physician delineated PTV(s) and 2 pseudo OARs: a 2-cm ring structure that extended from 7 mm to 27 mm outside of the union of all PTVs and all remaining tissue beyond the ring structure. This plan was optimized to conformally cover each PTV (plans had between 1 and 3 PTVs) with prescribed doses at the 95% level using a consistent in-house auto-planning algorithm developed in Pinnacle 16.2. The pseudo-OARs were used to encourage plan conformity with optimization objectives of maximum dose $<35$ Gy for the ring structure and 10% volume $<13$ Gy for all remaining tissue.

### OAR estimation

A standard set of head and neck OARs ($OAR_{AD}$) was autogenerated for each case using Pinnacle's model-based AD algorithm [Smart Probabilistic Image Contouring Engine (SPICE)].[6] This set contained 19 OARs that were used in the present analysis. Notably, this algorithm did not delineate the esophagus, larynx, or brachial plexus, which were clinically delineated in several cases. These structures were thus always designated as "high-priority" OARs in this study. Geometric agreement between AD and MD OARs was evaluated by computing the Dice overlap coefficient and the Hausdorff surface distance (HD) metric at the 50%, 95%, and maximum levels ($HD_{50}$, $HD_{95}$, $HD_{max}$).

### Conservative PQM estimation

The head and neck treatment plans evaluated used mean and maximum OAR dose metrics ($D_{mean}$ and $D_{max}$). $D_{mean}$ and $D_{max}$ were therefore conservatively estimated from $D_{est}$ and $OAR_{AD}$, while accounting for potential delineation errors, with the following process:

Let $OAR_{T,i}$ be the true delineation of $OAR_i$. Let $M$ be the maximum potential delineation error of auto-delineated $OAR_{AD,i}$ from $OAR_{T,i}$. That is, all points on

**Table 1** Patient characteristics

| Patient | Site | Type | Stage | Rx |
|---|---|---|---|---|
| 1 | Supraglottis | SCC | III | 70/56 Gy |
| 2 | BOT | SCC | III | 70/56 Gy |
| 3 | Soft palate | SCC | III | 70/56 Gy |
| 4 | Skull base | SCC | IVb | 70/56 Gy |
| 5 | Nasopharynx | SCC | III | 70/60/54 Gy |
| 6 | Tongue | SCC | III | 60/54 Gy |
| 7 | Tongue | SCC | IVa | 60/54 Gy |
| 8 | Oropharynx | SCC | IVb | 70/60/54 Gy |
| 9 | Lt. tonsil | SCC | IVa | 66/60/54 Gy |
| 10 | Rt. tonsil | SCC | IVa | 66/54 Gy |

*Abbreviations:* "BOT" = base of tongue; PTV = planning target volume; SCC = squamous cell carcinoma.
Prescriptions listed in terms of dose to PTV1, PTV2, and PTV3 (where applicable).

the surface of $OAR_{AD,i}$ must lie within $M$ mm of the $OAR_{T,i}$ surface. Although it is possible for $OAR_{AD,i}$ to underestimate $D_{max}$ for $OAR_{T,i}$, $D_{max}$ for $OAR_{T,i}$ is always less than or equal to the maximum dose in $OAR_{AD,i}$ expanded by an $M$ mm buffer. We thus expanded $OAR_{AD,i}$ by $M$ mm and conservatively assigned $D_{max}(OAR_{T,i})$ to be $D_{max}(OAR_{AD,i} + M)$.

Because uniform OAR expansions can extend into both high and low dose regions, the mean dose in $OAR_{AD,i} + M$ does not represent the worst case scenario $D_{mean}$ for $OAR_{T,i}$. The highest possible mean dose occurs when an OAR with minimal volume extends only into the highest dose region contained in the maximal-volume OAR envelope ($OAR_{AD,i} + M$). Assuming the surface of $OAR_{AD,i}$ lies within $R$ mm of the $OAR_{T,i}$ structure surface, the smallest possible structure can be generated by contracting $OAR_{AD,i}$ by $R$ mm: $OAR_{AD,i} - R$. The absolute volume of $OAR_{AD,i} - R$, $V_{min}$, can then be treated as a lower bound for the volume of $OAR_{T,i}$. Given a voxel volume, dV, this translates to a region containing $N_{min} = V_{min}/dV$ voxels. Because this region can also lie outside of $OAR_{AD,i}$, voxels in the expansion volume, $OAR_{AD,i} + M$, must also be considered. Thus, a conservative $D_{mean}$ estimate was computed by averaging the highest $N_{min}$ dose voxels contained in the $OAR_{AD,i} + M$ envelope. This process is demonstrated for 2 OARs in Figure 2.

For OARs that are small with respect to $R$, the contraction step can potentially reduce the volume to few or even zero voxels. To avoid this issue, cases in which $OAR_{AD,i} - R$ had a volume <10% of $OAR_{AD,i}$, $N_{min}$ was set to 10% of the $OAR_{AD,i}$ volume. Note that although, in general, the expansion ($M$) and contraction ($R$) buffers do not need to be equivalent, for simplicity we used $M = R$ in this study.

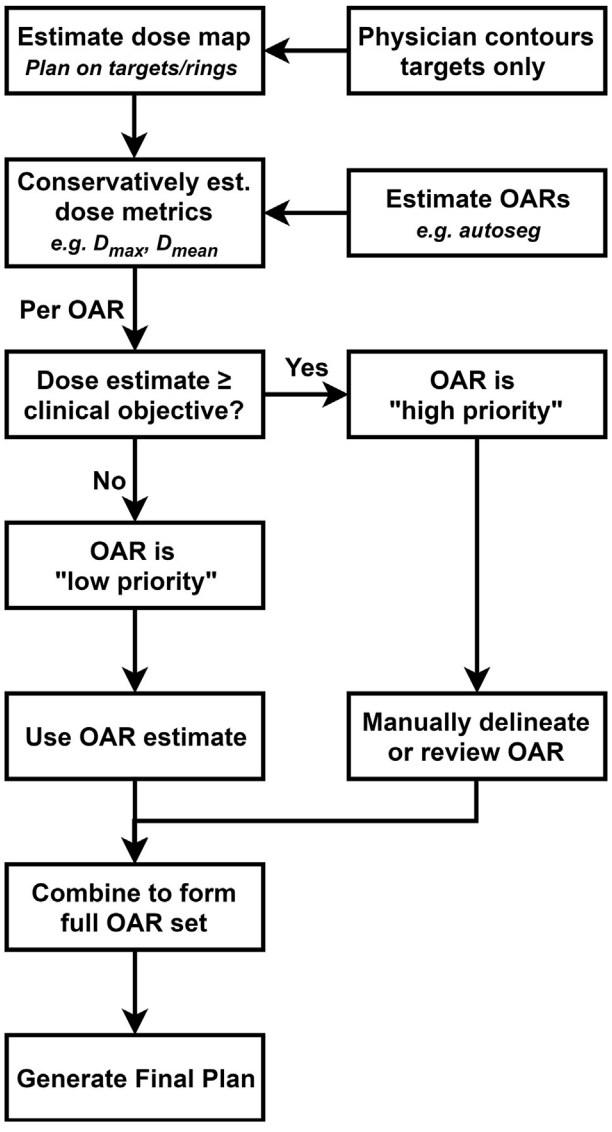


**Figure 1** Flowchart of the proposed method.

## OAR priority designation

OARs with conservatively estimated $D_{mean}$ or $D_{max}$ below their respective clinical objective values (Table 2) were deemed low priority and thus eligible for exclusion from MD or manual review. All other OARs were deemed high priority and in need of MD or manual review.

Because $M$ is defined as the maximum potential delineation error for an AD OAR, one might assume that the optimal choice of $M$ would be equal to the HD between AD and MD structures (ie, the maximum difference between structure surfaces). However, because the conservative dose maps used to estimate OAR PQM values differ from final clinical treatment plans, different values

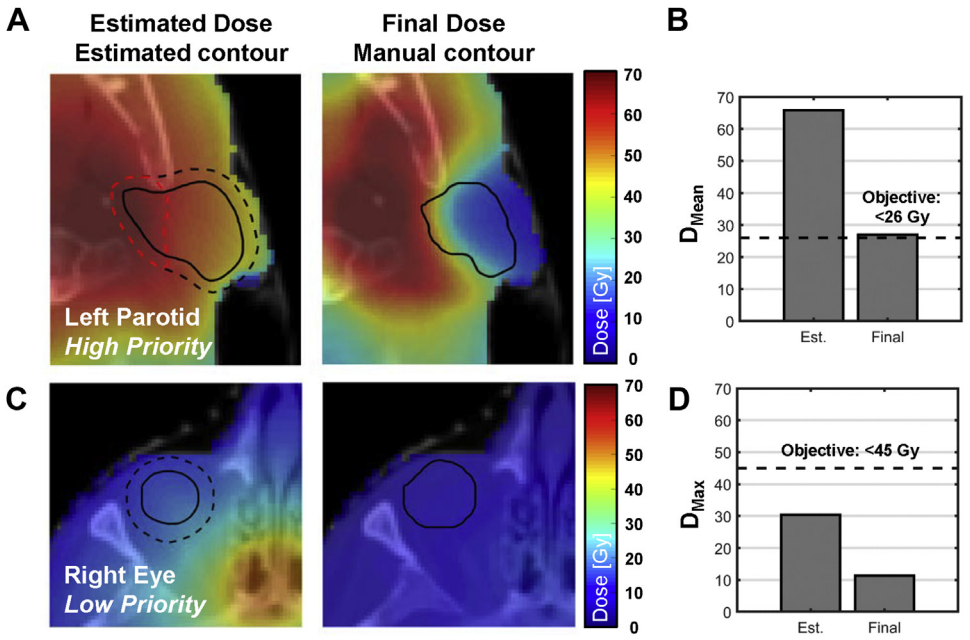


**Figure 2** Visualizations of the proposed method for a left parotid structure that was designated high priority (A, B) and a right eye structure that was designated low priority (C, D) with an uncertainty buffer M = 5 mm. Conservative dose estimates were computed using an initial dose map estimate (A, C). $D_{mean}$ estimates were computed in high dose subvolumes within auto-delineation (AD) structure expansions (ie, red dashed line in A), and $D_{max}$ estimates were computed in the entirety of AD structure expansions (ie, black dashed line in C). These estimates were compared with clinical objectives (black dashed lines in B,D) to determine organs at risk (OAR) priority levels. For final planning, low-priority AD structures were used without review and high-priority AD structures were replaced with manual delineation (MD) structures. Final plans were evaluated using all MD structures to evaluate clinical effect.

**Table 2** List of PQMs used as dosimetric OAR objectives in this study

| OAR | PQM | Dose |
|---|---|---|
| Larynx | $D_{mean}$ | 35 Gy |
| Esophagus | | 35 Gy |
| Pharyngeal constrictors | | 54 Gy |
| Parotid (rt. & lt.) | | 26 Gy |
| Cochlea (rt. & lt.) | | 35 Gy |
| Oral cavity | | 40 Gy |
| Brain | $D_{max}$ | 60 Gy |
| Brain stem | | 54 Gy |
| Optic nerve (rt. & lt.) | | 54 Gy |
| Chiasm | | 54 Gy |
| Spinal cord | | 45 Gy |
| Eye (rt. & lt.) | | 45 Gy |
| Mandible | | 70 Gy |
| Brachial plexus | | 70 Gy |

*Abbreviations:* OAR = organs at risk; PQM = plan quality metrics. PQMs were kept consistent between patients except when planning directives specified alternative PQMs (there were varying PQMs for larynx and cochlea). Note that not all OARs were present in all cases.

may be optimal for the task of identifying high- and low-priority structures. Therefore, the prioritization procedure was repeated for a range of 6 *M* values (0, 2.5 mm, 5 mm, 10 mm, 20 mm, and 40 mm) to identify the technique's sensitivity to this parameter and identify the value that minimizes manual contour review without violating OAR dose tolerances.

## Dosimetric evaluation

To determine the adequacy and clinical viability of OAR priority designations, 4 treatment plans were generated for each patient using different OAR structure sets for optimization: plan MD, all clinically used manually delineated OARs ($OAR_{MD}$); plan AD, all auto-delineated OARs ($OAR_{AD}$) plus manual OARs not included in the AD set (esophagus, larynx, and brachial plexus); plan mix, a mixed OAR set with high-priority OARs from $OAR_{MD}$ and low-priority OARs from $OAR_{AD}$; and plan $AD_{HP}$, only the high-priority OARs from $OAR_{MD}$ (low-priority OARs excluded).

All plans used 2 360° coplanar volumetric modulated arc therapy arcs and were optimized using consistent target volumes and a consistent auto-planning algorithm developed in Pinnacle 16.2. In this algorithm, plans were optimized to meet clinical objective values for each OAR (Table 2) in an iterative process. Optimization objectives were first initialized with values ~10 Gy below clinical objectives and an initial optimization run was performed. Each OAR was then evaluated in the resulting plan

according to its clinical objective value. Any OARs whose clinical objectives were not met at this stage had their optimization objectives reduced by 2 Gy to encourage further dose reduction before rerunning the optimization. This process was repeated twice for each case to generate a final treatment plan. The planning algorithm was uniform across all subjects. The PTV prescription doses were patient specific (Table 1).

After optimization, OAR PQMs were evaluated for all plans using $OAR_{MD}$ (ie, $OAR_{MD}$ was treated as ground-truth, regardless of what was used for planning). PQMs for each OAR and plan were then compared with clinical objective values and values from plan MD (ie, planned and evaluated using $OAR_{MD}$). Failed delineation priority designations were identified by OARs that were initially deemed low priority but, when used for optimization, resulted in plans that did not meet clinical objectives. PQM values were also compared directly between plans by computing the differences in $D_{mean}$ or $D_{max}$ between each plan and plan MD (ΔPQM). A 1-sample $t$ test was performed on ΔPQM across all OARs to test whether it significantly differed from 0.

Maximum point doses (ie, hotspots) and dose conformity index (volume of the prescription isodose level divided by PTV volume) were also compared between plans (ie, between plan MD, AD, mix, and $AD_{HP}$) to ensure consistent plan quality across optimizations. Each parameter was compared between plans across all subjects using 1-way analysis of variance.

## Results

### OAR priority designation

The 10-patient data set contained a total of 201 OARs that were contoured clinically. Of these, 168 were auto-delineated with SPICE. The average Dice overlap between AD and MD OARs was $0.59 \pm 0.29$. Average HD metrics were $HD_{50} = 3.4 \pm 2.4$ mm, $HD_{95} = 13.3 \pm 15.6$ mm, $HD_{max} = 18.3 \pm 18.5$ mm. OAR-specific agreement metrics are shown in Table 3. The number of OARs that were deemed low priority varied between patients and with the choice of delineation uncertainty parameter $M$ (Fig 3). When no buffer was used ($M = 0$), a total of 67 OARs were deemed low priority, representing a 33.3% reduction in the number of OARs requiring MD or manual review. This reduced to 54 low-priority OARs (26.9% reduction) with $M = 2.5$ mm, 49 with $M = 5$ mm (24.4% reduction), 43 with $M = 10$ mm (21.4% reduction), 32 with $M = 20$ mm (15.9% reduction), and 12 with $M = 40$ mm (6.0% reduction). In comparison, using all AD structures without modification reduced the number of manual delineations by 168 (83.6% reduction). Patient-specific low-priority designations and contouring reductions are shown in Table 4.

Specific examples for OARs in a nasopharyngeal case are shown in Figure 2. In this case (shown with $M = 5$ mm), the mean left parotid dose was conservatively estimated (within the red dotted area) to exceed the clinical objective value of 26 Gy and thus was identified as a high-priority structure (Fig 2A, B). Analysis of the final plan using an MD left parotid structure showed a final $D_{mean}$ close to the objective value and a high-dose gradient near contour edges, indicating that the structure contributed appreciably to the optimization objective and result (and thus a different OAR delineation could have resulted in a significantly different plan). In contrast, the right eye maximum dose was conservatively estimated (within the black dotted area) to be below the clinical objective value of 45 Gy and thus was considered low priority (Fig 2C, D). The final plan showed little dose gradient in the right eye area and a final $D_{max}$ well below the 45 Gy objective, indicating that the specific structure boundaries did not have a large effect on the plan. Note, however, that the final $D_{max}$ did reduce considerably between the initial and final dose maps (though both were substantially below 45 Gy). This is due to the presence of other OARs that were deemed high priority (eg, the oral cavity and right cochlea) between the target volume and right eye, which indirectly reduced dose to this area.

### Dosimetric evaluation

No significant differences in dose hotspot ($1.14 \pm 0.07$, $P > .99$) or conformity index ($1.26 \pm 0.21$, $P > .99$) were observed across plans, indicating a consistent auto-planning technique. The dosimetric effect of the OAR prioritization workflow is demonstrated in Figures 4 and 5. In each of these figures, the achieved PQM (ie, $D_{mean}$ or $D_{max}$) for each OAR (computed using $OAR_{MD}$) is plotted as a percentage difference from its clinical objective value (ΔPQM) for plan MD in the x-axis and plans AD, mix, (Fig 4), and $AD_{HP}$ (Fig 5) in the y-axis. In these plots, the lower-left quadrant contains OARs that met clinical objectives in both plan MD and the indicated alternative plan, whereas the upper right quadrant indicates objective failures in both cases. Similarly, the upper left quadrant indicates OARs that met objectives in plan MD but failed in the alternative plan. Points highlighted in red indicate structures determined to be low priority at the specified $M$ setting. As $M$ increased, the structures that met low-priority criteria became limited to those further from clinical tolerance levels (ie, toward the lower left corners of the scatter plots). Interestingly, a few low-priority points did remain, with doses close to tolerance levels (ie, toward the plot centers), even with large values of $M$. These points all represent OARs with $D_{max}$ objectives that approach prescription dose levels. For example, for $M = 40$ mm, all 3 of the low-priority points that are observed near the center of the plot are

**Table 3** Geometric agreement indices between manually delineated and auto-delineated OARs

| OAR | Dice | $HD_{50}$ (mm) | $HD_{95}$ (mm) | $HD_{max}$ (mm) |
|---|---|---|---|---|
| Brain | 0.97 ± 0.0 | 4.4 ± 0.7 | 11.2 ± 1.7 | 32.2 ± 3.2 |
| Brain stem | 0.83 ± 0.0 | 2.0 ± 0.2 | 5.1 ± 1.1 | 8.5 ± 2.0 |
| Chiasm | 0.19 ± 0.2 | 6.2 ± 2.7 | 11.8 ± 4.7 | 12.7 ± 4.4 |
| Cochleae | 0.20 ± 0.1 | 4.6 ± 1.6 | 8.8 ± 2.0 | 9.2 ± 2.0 |
| Eyes | 0.85 ± 0.0 | 1.7 ± 0.2 | 4.2 ± 1.1 | 6.0 ± 1.4 |
| Glottis | 0.49 ± 0.1 | 3.7 ± 0.2 | 8.0 ± 2.4 | 9.8 ± 2.5 |
| Lenses | 0.50 ± 0.2 | 2.3 ± 1.4 | 4.3 ± 1.6 | 4.7 ± 1.7 |
| Mandible | 0.86 ± 0.0 | 1.3 ± 0.1 | 5.9 ± 4.8 | 13.0 ± 7.1 |
| Optic nerves | 0.47 ± 0.1 | 2.3 ± 0.7 | 7.1 ± 2.9 | 7.9 ± 3.2 |
| Oral cavity | 0.57 ± 0.1 | 8.3 ± 4.6 | 21.3 ± 4.7 | 25.7 ± 5.1 |
| Parotids | 0.70 ± 0.1 | 3.1 ± 0.8 | 11.8 ± 5.0 | 16.8 ± 6.8 |
| Pharynx const | 0.09 ± 0.2 | 3.3 ± 1.3 | 24.1 ± 14.7 | 30.4 ± 15.9 |
| Spinal cord | 0.69 ± 0.1 | 3.3 ± 2.0 | 49.5 ± 38.9 | 61.9 ± 38.6 |

*Abbreviations:* HD = Hausdorff distance; OAR = organs at risk.

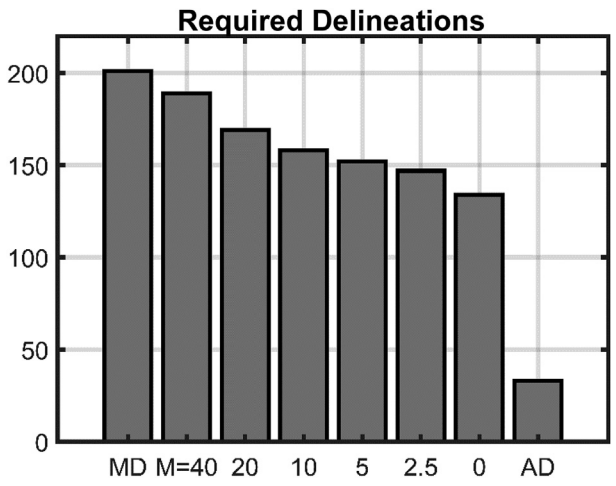


**Figure 3** The number of manual delineations (MDs) required across 10 head and neck patients when using only MD, the proposed organs at risk (OAR) prioritization strategy with varying delineation uncertainty parameters *(M)*, and using auto-delineation (AD) for all available structures.

mandible contours with 70 Gy $D_{max}$ objectives. The other OARs that approach the center for $M = 10$ mm and 20 mm are brain ($D_{max}$ objective: 60 Gy) and brain stem ($D_{max}$ objective: 54 Gy). Because these plots are expressed as percentages of objective values, a substantial dose deviation may appear small (ie, a brain stem $D_{max}$ 14 Gy below the 54 Gy objective equates to $\Delta PQM = -26\%$). Additionally, when objectives approach the prescription doses themselves (ie, mandible), OARs may never exceed tolerances, despite large expansions that closely approach targets, as long as the plan does not have a hotspot in that area.

When all AD OARs, regardless of priority, were used during optimization (plan AD), 7 structures that passed clinical objectives in plan MD ($\Delta PQM < 0$) exceeded their objectives in plan AD, indicating clinically meaningful variations (ie, points in the upper left quadrant in Fig 4). Three of these failures were parotid glands, 2 were spinal cords, 1 was cochlea, and 1 a lens. Notably, these structures were not outliers in terms of geometric agreement with MD (average Dice = 0.67 among these 7 vs 0.59 overall) and were all deemed high priority for all uncertainty settings (including $M = 0$).

When OAR prioritization was used (with any value of *M*) and low-priority AD structures plus high-priority MD structures were used for optimization (ie, plan mix), no objectives that were met in plan MD failed (Fig 4). One OAR (a brain) that was deemed low priority with $M = 0$ did exceed its objective in all plans (MD, AD, mix, and $AD_{HP}$), indicating that the proposed workflow understated the final OAR dose in this case. However, with any $M > 0$, no low-priority OARs exceeded objective values in plan MD or plan mix, indicating that with the additional buffer for delineation uncertainty, low-priority designations were always appropriate.

When low-priority OARs were ignored rather than included using AD structures (plan $AD_{HP}$), between 1 and 3 OARs (depending on *M* ) exceeded objectives that were met in plan MD (Fig 5). This indicates that in all but a few cases, low-priority OARs could be completely ignored without consequence, but, in these few cases, AD OAR objectives were required to guide the optimization to a clinically acceptable dose distribution.

Absolute PQM values (ie, $D_{mean}$ and $D_{max}$) did not increase significantly between plan MD and any of the alternative plans (AD, $AD_{HP}$, and mix) for any value of *M*. For hybrid contour plans (plan mix) the following ΔPQMs were observed across all OARs: $M = 0$, $\Delta PQM = 0.32 \pm 6.28\%$ ($P = .51$); $M = 2.5$ mm, $\Delta PQM = -0.36 \pm 7.25\%$ ($P = .52$); $M = 5$ mm, $\Delta PQM = 0.32 \pm 6.28\%$ ($P = .51$); $M = 10$ mm, $\Delta PQM = -0.42 \pm 6.13\%$ ($P = .38$); $M = 20$ mm, $\Delta PQM = -0.28 \pm 6.45\%$ ($P = .57$); $M = 40$ mm, $\Delta PQM = 0.18 \pm 6.76\%$ ($P = .73$). Additionally, no significant increases were found across specifically $D_{mean}$ or $D_{max}$ objective OARs.

## Discussion

The proposed workflow reduced the number of OARs requiring manual delineation or review by 33% across 10 patients with head and neck cancers and up to 63% for individual patients without affecting any clinically relevant OAR doses. Importantly, this was accomplished using auto-contours and dose map predictions generated using commercially available software and thus can be

**Table 4** Per-patient and total OAR reductions with each level of assumed DV ($M$)

| Site | Total OARs | Number of low priority OARs (% of total) | | | | | |
|---|---|---|---|---|---|---|---|
| | | $M = 0$ | $M = 2.5$ mm | $M = 5$ mm | $M = 10$ mm | $M = 20$ mm | $M = 40$ mm |
| Supraglottis | 23 | 12 (52.2) | 11 (47.8) | 10 (43.5) | 9 (39.1) | 8 (34.8) | 7 (30.4) |
| BOT | 17 | 5 (29.4) | 4 (23.5) | 4 (23.5) | 4 (23.5) | 4 (23.5) | 1 (5.9) |
| Soft palate | 28 | 7 (25.0) | 5 (17.9) | 5 (17.9) | 5 (17.9) | 5 (17.9) | 0 (0.0) |
| Skull base | 17 | 3 (17.6) | 3 (17.6) | 2 (11.8) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Nasopharynx | 24 | 4 (16.7) | 3 (12.5) | 2 (8.3) | 2 (8.3) | 0 (0.0) | 0 (0.0) |
| Tongue | 16 | 10 (62.5) | 6 (37.5) | 6 (37.5) | 6 (37.5) | 5 (31.3) | 2 (12.5) |
| Tongue | 14 | 6 (42.9) | 5 (35.7) | 5 (35.7) | 4 (28.6) | 4 (28.6) | 1 (7.1) |
| Oropharynx | 22 | 10 (45.5) | 8 (36.4) | 6 (27.3) | 5 (22.7) | 2 (9.1) | 0 (0.0) |
| Lt. tonsil | 20 | 4 (20.0) | 3 (15.0) | 3 (15.0) | 2 (10.0) | 1 (5.0) | 1 (5.0) |
| Rt. tonsil | 20 | 6 (30.0) | 6 (30.0) | 6 (30.0) | 6 (30.0) | 3 (15.0) | 0 (0.0) |
| Total | 201 | 67 (33.3) | 54 (26.9) | 49 (24.4) | 43 (21.4) | 32 (15.9) | 12 (6.0) |

*Abbreviations:* BOT = base of tongue; DV = delineation variability; OARs = organs at risk.

easily implemented clinically. Furthermore, this workflow can be fully automated to require no additional manual effort. Once target volumes and prescriptions are defined by a physician, the AD and dose map estimation processes can be initiated such that priority designations are available when OAR contouring and treatment planning begins.

Decisions on which structures to contour or not contour in a given plan currently rely on subjective and ad hoc analysis. Although this may be adequate in obvious situations (eg, omitting pelvic structures in a head and neck plan), it is unlikely to be reliable in borderline scenarios and can result in over contouring. By quantifying this intuitive logic, our approach can make objective decisions in nonobvious situations and reduce unnecessary contouring that would not otherwise be identified. This will allow planners to both increase their efficiency and improve the quality of their treatment plans by reducing the time spent on clinically unimportant tasks (eg, reviewing low-priority structures). Furthermore, by explicitly identifying structures that do require review, this method may improve compliance in thorough-structure review and allow planners to spend more time on high-priority structures that will consequently improve the quality of important delineations.

The dosimetric equivalence of plans generated using low-priority AD OARs and high-priority MD OARs for optimization with reference plans (Fig 4) indicates that our algorithm correctly identified high- and low-priority OARs. Although in most cases, even ignoring low-priority OARs led to equivalent plan quality (Fig 5), this did lead to up to 3 failing OAR objectives that were met in reference plans. This indicates that in most cases, low-priority OARs were sufficiently far from objective doses that they were not at risk of exceeding limits even if excluded entirely from optimization, but that in some cases, the inclusion of an AD structure was still necessary.

The choice of uncertainty parameter $M$ did not affect clinically significant dose metrics, but did have a strong effect on the number of OARs deemed low priority by our algorithm (and thus the associated efficiency gains, Fig 3). $M$ describes the size of the buffer used to conservatively estimate $D_{mean}$ or $D_{max}$ from AD structures and thus should, in principle, reflect the confidence that one has in the AD strategy being used. However, even the most aggressive setting ($M = 0$, ie, no buffer) resulted in zero objective failures among low-priority OARs despite imperfect agreement between AD and MD structures (mean Dice = 0.59). This indicates that in the present implementation, $M = 0$ provides optimal results (ie, maximum efficiency with no clinical effect). However, this is likely due to the highly conservative nature of our dose map estimates, which did not attempt to predict any conformal OAR avoidance, which likely balanced out the delineation differences between MD and AD OARs. Use of a different AD algorithm or different treatment site might yield a different optimal $M$.

Several other techniques have been developed to estimate OAR doses before planning by predicting dose volume histograms (ie, knowledge based planning)[7-12] and/or full 3-dimensional dose maps.[13,14] Any of these methods could be used to provide initial OAR dose estimates for our algorithm. However, it is worth noting that these methods generally aim to predict achievable dosimetry, whereas our method aims to conservatively estimate potential OAR doses. Although these methods might more accurately predict final dose distributions than our approach, they would likely make the OAR prioritization workflow more sensitive to contouring variability and necessitate larger buffers ($M$) to achieve optimal performance. Dosimetric buffers could also be used to account for uncertainty in dose prediction (ie, comparing OAR doses with a dose criterion below the desired objective value) in addition to or in place of geometric buffers.

Pinnacle's model-guided auto-contouring algorithm (SPICE)[6] generally produced OARs that were in reasonable agreement with manual delineations. However,

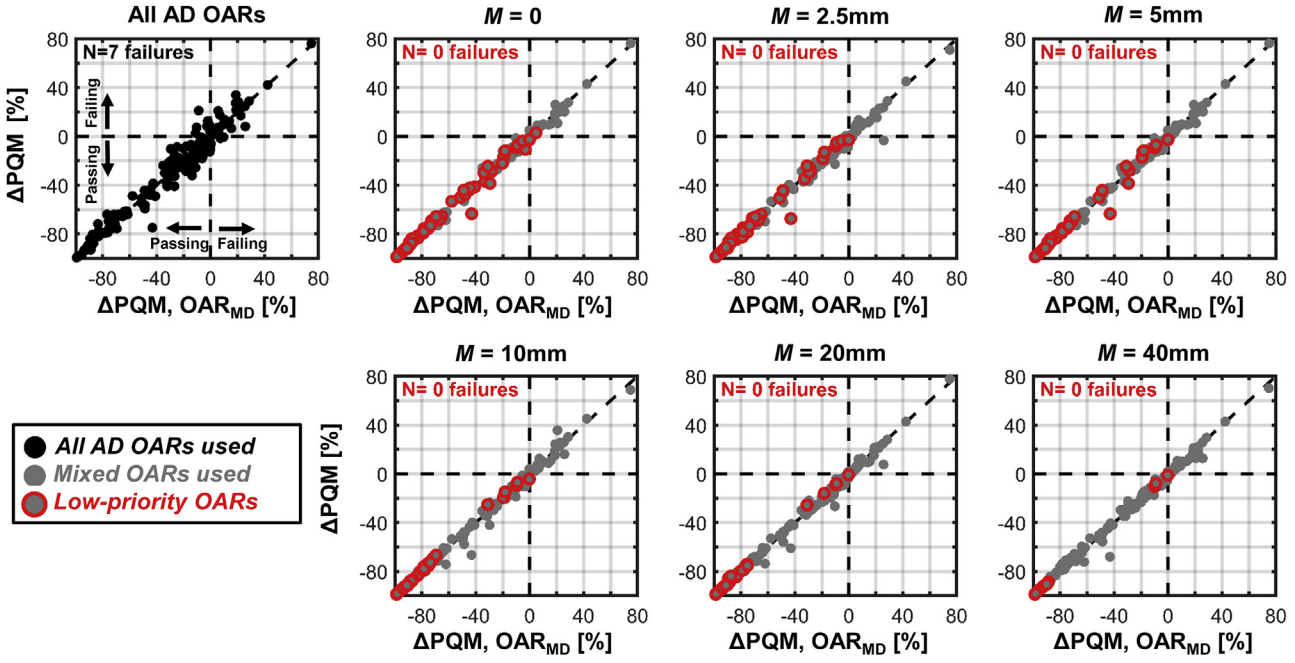


**Figure 4** Percent differences between achieved organs at risk (OAR) plan quality metrics (PQMs) ($D_{mean}$ or $D_{max}$) and objectives (ΔPQM) from reference plans that used manual delineation (MD) OARs for optimization (x-axis, plan MD) and test plans that incorporated auto-delineation (AD) OARs (y-axis, plan AD: black dots or plan mix: gray dots). Negative ΔPQM values indicate OAR PQMs below objectives and positive values indicate failed objectives.

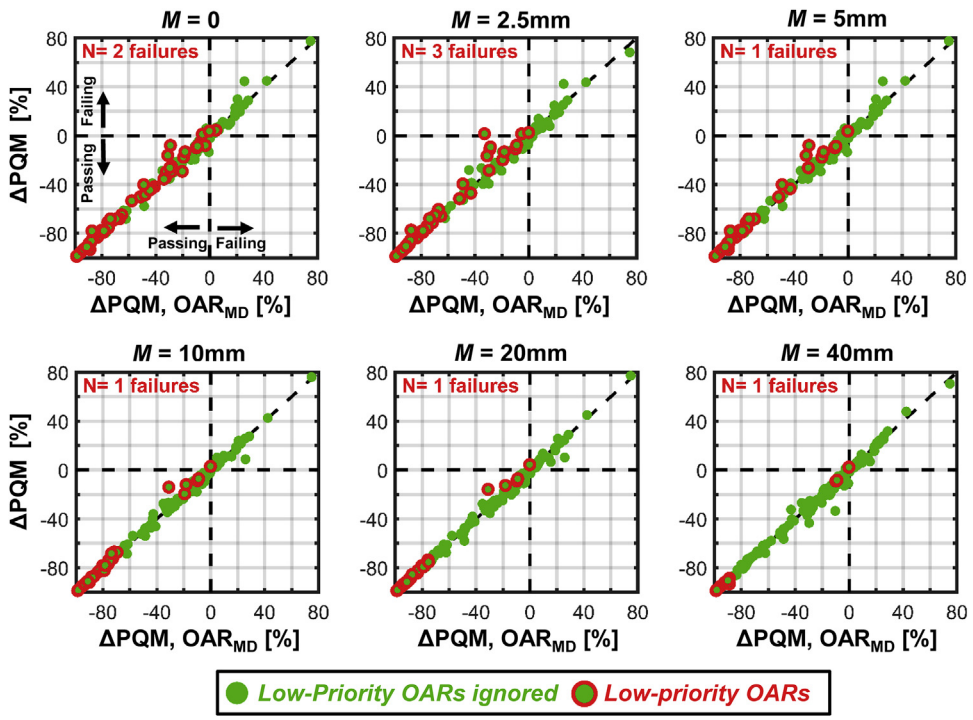


**Figure 5** Scatter plots analogous to those in Figure 4 showing percent differences from plan quality metrics (PQM) objectives (ΔPQM) achieved from plans generated using only high-priority auto-delineation (AD) organs at risk (OAR) for planning (y-axis, plan $AD_{HP}$) as functions of baseline values from plans generated using all manual delineation (MD) OARs (x-axis, plan MD). Numbers in the upper left corners indicate the number of OARs that passed their objectives in plan MD but failed in plan $AD_{HP}$.

simply using all (unmodified) AD OARs in plan optimizations resulted in 7 failing objectives that were met when MD OARs were used (Fig 4). Although this is only a small portion (3.5%) of the 201 OARs examined, this clearly indicates that not all AD OARs could safely be omitted from manual review. However, this also shows that for the vast majority of OARs (96.5%), manually tweaking or redrawing would not have been clinically consequential. This underscores the importance of methods that can identify specifically which OARs will benefit from manual modification to both ensure plan quality and efficient use of human attention.

The proposed workflow is also fully compatible with other auto-delineation methods and could therefore benefit from the rapid advancements that are occurring in that area. For example, state-of-the-art machine-learning-based AD methods[15-22] could generate structures that more closely agree with MD volumes and improve our method's ability to prioritize OARs. As auto-delineation accuracy improves, more conformal dose estimates can also be used to estimate OAR doses without the need for large spatial buffers. However, given a perfect AD technique (ie, one with zero disagreement with MD), all structures could in principle be excluded from review with no effect on dosimetry. The imperfect AD technique used in this work was thus a valuable test case for our method.

Minimizing the time spent on OAR delineation and review is particularly important in adaptive RT due to the need for new delineations for each treatment fraction. AD is routinely used in this context and is typically accomplished using deformable image registration to propagate contours from a previous CT or magnetic resonance imaging onto a daily image.[23-27] Manual or programmatic supervision of this process is generally required to ensure the accuracy of contour propagations.[28] Our OAR prioritization workflow can be used in concert with such quality assurance measures to reduce the number of structures requiring manual attention while a patient is on the treatment table. In this context, existing dose maps and achieved OAR dose metrics from previous fractions can be used to more accurately initialize the prioritization process. Further efficiency gains could also be achieved within our methodology by identifying specific high-priority slices or regions within OARs and/or providing a graphical indication of OAR-specific delineation accuracy requirements.

One related approach that has been proposed for magnetic resonance imaging–guided adaptive RT in the abdomen is to only review and/or modify deformably mapped OARs in regions that lie within 3 cm of the PTV surface.[29] This method is effective for prioritizing OARs with $D_{max}$ objectives that are close to prescription dose levels but cannot accurately account for $D_{mean}$ objectives or OARs that are sensitive to lower dose levels. Our method overcomes these limitations by using dosimetric rather than geometric criteria.

The present study did have some limitations that should be discussed. First, a key assumption in our workflow is that the initial dose estimate delivers more dose to all OARs than will be delivered by the final plan. This was a reasonable assumption for our particular dose prediction strategy because it did not attempt to model conformal OAR avoidance, but, as discussed previously, may not hold in future implementations with more sophisticated dose prediction strategies. This issue could be addressed with an additional loop in the workflow wherein OAR priority level is reassessed after a final plan is generated and previously low-priority OARs may be recommended for MD or manual review. We also relied on an assumption that AD OARs were free of gross errors such as placement in a completely disparate region of the body. We did not encounter such errors in the present study but, if present, they could confound our algorithm and potentially misidentify low-priority OARs. Clinically validated AD algorithms or automated OAR quality control programs[30] should therefore be used to mitigate this risk. Furthermore, although the present method relied on simple OAR expansions and contractions to account for delineation uncertainties, additional steps could be taken to refine these operations in making conservative dose estimates. For example, expansions of soft-tissue structures into bone could be automatically cropped out, which would likely improve performance.

Finally, although low-priority OARs do not meaningfully affect plan quality, their inaccurate delineation could result in inaccuracies in dose volume histogram and PQM reporting if not indicated properly. Although these inaccuracies would not be clinically meaningful, it is important that nonreviewed low-priority AD structures be clearly labeled as such in the treatment plan to ensure clear reporting. This would emphasize low-priority structures that were not clinically relevant in the current treatment plan but must be reviewed for any subsequent usage. For example, if a patient is retreated, previously low-priority structures may become high priority in the context of their new treatment plan and therefore must be reviewed before clinical decisions are made. This labeling would also indicate that these structures must be reviewed before their use in clinical research. Alternatively, one could choose to simply exclude nonreviewed low-priority structures from the patient chart.

## Conclusions

We have presented an automated workflow that identifies OARs that do and do not require manual attention before treatment planning, facilitating the safe and efficient clinical integration of auto-delineations. Our method correctly identified OARs that were likely to benefit from manual review and/or modification before planning and

permitted a 33% reduction in manual effort without affecting clinically relevant OAR doses.

## References


1. Sharp G, Fritscher KD, Pekar V, et al. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. *Med Phys*. 2014; 41:050902.
2. Lustberg T, van Soest J, Gooding M, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiother Oncol*. 2018;126:312-317.
3. Reed VK, Woodward WA, Zhang L, et al. Automatic segmentation of whole breast using atlas approach and deformable image registration. *Int J Radiat Oncol*. 2009;73:1493-1500.
4. Hoang Duc AK, Eminowicz G, Mendes R, et al. Validation of clinical acceptability of an atlas-based segmentation algorithm for the delineation of organs at risk in head and neck cancer. *Med Phys*. 2015;42:5027-5034.
5. Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. *Radiother Oncol*. 2016;121: 169-179.
6. Qazi AA, Pekar V, Kim J, Xie J, Breen SL, Jaffray DA. Auto-segmentation of normal and target structures in head and neck CT images: A feature-driven model-based approach. *Med Phys*. 2011; 38:6160-6170.
7. Wu B, Ricchetti F, Sanguineti G, et al. Data-driven approach to generating achievable dose–volume histogram objectives in intensity-modulated radiotherapy planning. *Int J Radiat Oncol*. 2011;79:1241-1247.
8. Moore KL, Brame RS, Low DA, Mutic S. Experience-based quality control of clinical intensity-modulated radiotherapy planning. *Int J Radiat Oncol*. 2011;81:545-551.
9. Appenzoller LM, Michalski JM, Thorstad WL, Mutic S, Moore KL. Predicting dose-volume histograms for organs-at-risk in IMRT planning. *Med Phys*. 2012;39:7446-7461.
10. Tol JP, Delaney AR, Dahele M, Slotman BJ, Verbakel WFAR. Evaluation of a knowledge-based planning solution for head and neck cancer. *Int J Radiat Oncol*. 2015;91:612-620.
11. Tran A, Woods K, Nguyen D, et al. Predicting liver SBRT eligibility and plan quality for VMAT and 4π plans. *Radiat Oncol*. 2017; 12:70.
12. Ahmed S, Nelms B, Gintz D, et al. A method for *a priori* estimation of best feasible DVH for organs-at-risk: Validation for head and neck VMAT planning. *Med Phys*. 2017;44:5486-5497.
13. Nguyen D, Jia X, Sher D, et al. 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture. *Phys Med Biol*. 2019;64:065020.
14. Shiraishi S, Moore KL. Knowledge-based prediction of three-dimensional dose distributions for external beam radiotherapy. *Med Phys*. 2015;43:378-387.
15. van Rooij W, Dahele M, Ribeiro Brandao H, Delaney AR, Slotman BJ, Verbakel WF. Deep learning-based delineation of head and neck organs at risk: Geometric and dosimetric evaluation. *Int J Radiat Oncol*. 2019;104:677-684.
16. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: Learning dense volumetric segmentation from sparse annotation. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Cham, Switzerland: Springer Nature; 2016:424-432.
17. Fritscher K, Raudaschl P, Zaffino P, Spadea MF, Sharp GC, Schubert R. Deep neural networks for fast segmentation of 3D medical images. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Cham, Switzerland: Springer Nature; 2016:158-165.
18. Tong N, Gou S, Yang S, Ruan D, Sheng K. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. *Med Phys*. 2018;45:4558-4567.
19. Ren X, Xiang L, Nie D, et al. Interleaved 3D-CNNs for joint segmentation of small-volume structures in head and neck CT images. *Med Phys*. 2018;45:2063-2075.
20. Nikolov S, Blackwell S, Mendes R, et al. Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. *arxiv*. 2018:1809.04430.
21. Zhu W, Huang Y, Zeng L, et al. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med Phys*. 2019;46:576-589.
22. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med Phys*. 2017;44:547-557.
23. Zhang T, Chi Y, Meldolesi E, Yan D. Automatic delineation of on-line head-and-neck computed tomography images: Toward on-line adaptive radiotherapy. *Int J Radiat Oncol*. 2007;68: 522-530.
24. Kumarasiri A, Siddiqui F, Liu C, et al. Deformable image registration based automatic CT-to-CT contour propagation for head and neck adaptive radiotherapy in the routine clinical setting. *Med Phys*. 2014;41:121712.
25. Hou J, Guerrero M, Chen W, D'Souza WD. Deformable planning CT to cone-beam CT image registration in head-and-neck cancer. *Med Phys*. 2011;38:2088-2094.
26. Lamb J, Cao M, Kishan A, et al. Online adaptive radiation therapy: Implementation of a new process of care. *Cureus*. 2017;9.
27. Lim-Reinders S, Keller BM, Al-Ward S, Sahgal A, Kim A. Online adaptive radiation therapy. *Int J Radiat Oncol Biol Phys*. 2017;99: 994-1003.
28. Beasley WJ, McWilliam A, Slevin NJ, Mackay RI, van Herk M. An automated workflow for patient-specific quality control of contour propagation. *Phys Med Biol*. 2016;61:8577-8586.
29. Bohoudi O, Bruynzeel AME, Senan S, et al. Fast and robust online adaptive planning in stereotactic MR-guided adaptive radiation therapy (SMART) for pancreatic cancer. *Radiother Oncol*. 2017; 125:439-444.
30. Hui CB, Nourzadeh H, Watkins WT, et al. Quality assurance tool for organ at risk delineation in radiation therapy using a parametric statistical approach. *Med Phys*. 2018;45:2089-2096.