Supplementary Material

Patient cohort and treatment planning

We considered all clival chordoma and chondrosarcoma patients treated at our institute between 2013 and 2015, with readily available data pertinent to the study. Since clival chordoma and chondrosarcoma are extremely rare diseases, the number of selected patients was small: 41 treated with passively scattered proton therapy (PSPT) and 10 treated with pencil beam scanning (PBS) proton therapy. This cohort received a combination of proton beam therapy (PBT) and intensity modulated radiation therapy (IMRT). The two modalities have the same clinical target. The number of IMRT fractions (usually 10) provide some flexibility in machine scheduling during the long course of treatment.

Chondrosarcoma patients are prescribed 70 Gy at 2 Gy per fraction. Chordoma patients are prescribed 76 Gy at 2 Gy per fraction, which is reduced to 68.4 Gy at 1.8 Gy per fraction if the patient exhibits certain clinical factors (e.g. diabetes, brainstem injury from surgery). If a conflict arises between target coverage and organ-at-risk (OAR) sparing such that target coverage falls below 50%, then the prescribed dose is lowered to 72 Gy at 2 Gy per fraction. Dose constraints to OARs are displayed in Table 1. Note that some exceptions are made to planning objectives in order to account for patient-specific clinical factors. The median (range) prescribed dose was 72 Gy (50.4 – 82 Gy) for the combined therapy, 52 Gy (30 – 62 Gy) for PBT, and 20 Gy (10 – 30 Gy) for IMRT. Personalized dose constraints may also be applied to OARs. For example, if a patient is already deaf in one ear, then particular care is taken to spare the contralateral cochlea.

Organ-at-risk	**Objectives**
Brainstem	Max dose 67 Gy (surface) and 55 Gy (center)
Spinal cord	Max dose 67 Gy (surface) and 55 Gy (center)
Optic nerve & chiasm	Max dose 62 Gy
Cochlea	As low as possible; max dose 50 Gy
Parotid gland	As low as possible; mean dose < 26 Gy

Table 1. Treatment planning objectives for normal tissues.

IMRT treatment planning targeted the planning target volume (PTV), which was created using a conventional expansion of the clinical target volume (CTV). In contrast, PBT treatment planning targeted the CTV and provided distal and lateral margins by adjusting the range, modulation width and aperture used by each beam. This yielded a distal margin of 3.5% of the range plus an additional 1 mm and a lateral margin of 8-10 mm (Paganetti, 2012). This is similar to using beam-specific asymmetric CTV expansions.

Knowledge-based method

The method for predicting organ-at-risk (OAR) dose-volume histograms (DVHs) is both geometric and knowledge-based in nature. It fundamentally assumes that the DVH of an OAR subvolume containing a small range of distance-to-target is universal. Thus interpatient variation in OAR DVH originates solely from differences in the distance-to-target histogram. These universal spatial patterns are measured in existing treatment plans in order to train a model, which can then predict the OAR DVH for a new patient. Deviations from this

fundamental assumption lead to a spread in the model parameters during training and degradation of prediction accuracy.

The process of training a model for a specific OAR type is summarized in Figure 1. First, one identifies every OAR structure of that type in the training cohort, including symmetric pairs (i.e. left and right instances). Second, each OAR structure is split into a set of subvolumes according to some geometric selection criteria. In this work, the kth selection criteria is $(k-1)w < r \leq kw$, where r is the distance-to-target and $w = 3\ \mathrm{mm}$ is the shell width. These criteria select a set of shells surrounding the surface of the clinical target volume (CTV). It is noted that $k \leq 0$ corresponds to OAR subvolumes that overlap with the CTV structure.

The differential DVH within each subvolume is fit using a probability distribution function $f(D;\boldsymbol{\theta})$, where D is the dose variable and $\boldsymbol{\theta}$ are the parameters. Following Appenzoller et al (2012), we used a skew normal distribution for this purpose

$$f(x;\boldsymbol{\theta}) = \sqrt{\frac{2}{\pi\theta_2^2}}\, e^{-(x-\theta_1)^2/(2\theta_2^2)} \int_{-\infty}^{\theta_3(x-\theta_1)/\theta_2} e^{-t^2/2}\,\mathrm{d}t$$

whose parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, \theta_3\}$ correspond to location, scale and shape respectively. Fitting the kth subvolume of the ith OAR structure yields the maximum-likelihood estimates $\widehat{\boldsymbol{\theta}}_{ik}$. Fit convergence is improved by bounding parameters within the ranges: $-1 < \theta_1 < 2$, $10^{-9} < \theta_2 < 1$ and $-10 < \theta_3 < 10$. Subvolumes with fewer than 4 voxels are excluded from fitting, since they cannot sufficiently constrain the parameters $\boldsymbol{\theta}$. The subvolume differential DVH is binned according to the Freedman-Diaconis rule.

Finally, sample estimates of the population parameters $\boldsymbol{\theta}_k$ are acquired for each subvolume by averaging over the OAR structures

$$\widehat{\boldsymbol{\theta}}_k = \frac{\sum_{i=1}^{N} \widehat{\boldsymbol{\theta}}_{ik}}{N}$$

These subvolume parameter estimates $\widehat{\boldsymbol{\theta}}_k$ embody the model. Subvolumes found in only a single structure (i.e. $N = 1$) are excluded. The subvolume selection criteria are chosen such that the subvolume DVHs can be described by the $f(D;\boldsymbol{\theta})$ distribution, and the entire population is well described by the parameter estimates $\widehat{\boldsymbol{\theta}}_k$. The degree to which these two conditions are met determines the model accuracy.

The way in which the parameter estimates $\widehat{\boldsymbol{\theta}}_k$ are used to predict the DVH of a new OAR structure depends upon the size of the structure. For structures containing greater than 100 voxels, the prediction process resembles the training process (see Figure 2). The OAR structure is split into a set of subvolumes according to the same geometric selection criteria, in order to obtain fractional volumes v_k. These are used to weight the predicted subvolume DVHs such that the predicted total DVH is

$$\sum_k v_k f\left(D;\widehat{\boldsymbol{\theta}}_k\right)$$

For structures containing fewer than 100 voxels, it was found to be advantageous to predict the DVH of each voxel separately by interpolating parameter estimates $\widehat{\boldsymbol{\theta}}_k$ between subvolumes (see Figure 3). Minimal gains were observed when this method was applied to structures with greater than 100 voxels. The interpolation was simple to achieve in this

study because the distance-to-target was the sole quantity used to separate subvolumes, and therefore it was possible to interpolate $\widehat{\boldsymbol{\theta}}_k$ using a one-dimensional cubic smoothing spline (weighted by the inverse variance in $\widehat{\boldsymbol{\theta}}_{ik}$ and using a smoothing factor of 0.8). Parameters were extrapolated outside the fitted distance-to-target range by using the parameter estimates at the minimum and maximum distance-to-target used in the fitting. Then the predicted total DVH is

$$\sum_j f\left(D; \widehat{\boldsymbol{\theta}}(r_j)\right)$$

where j labels voxels.

References

1. Appenzoller LM, Michalski JM, Thorstad WL, et al. Predicting dose-volume histograms for organs-at-risk in IMRT planning. *Med Phys* 2012;39:7446-7461. http://dx.doi.org/10.1118/1.4761864
2. Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Phys Med Biol* 2012;57:R99-177. http://dx.doi.org/10.1088/0031-9155/57/11/R99

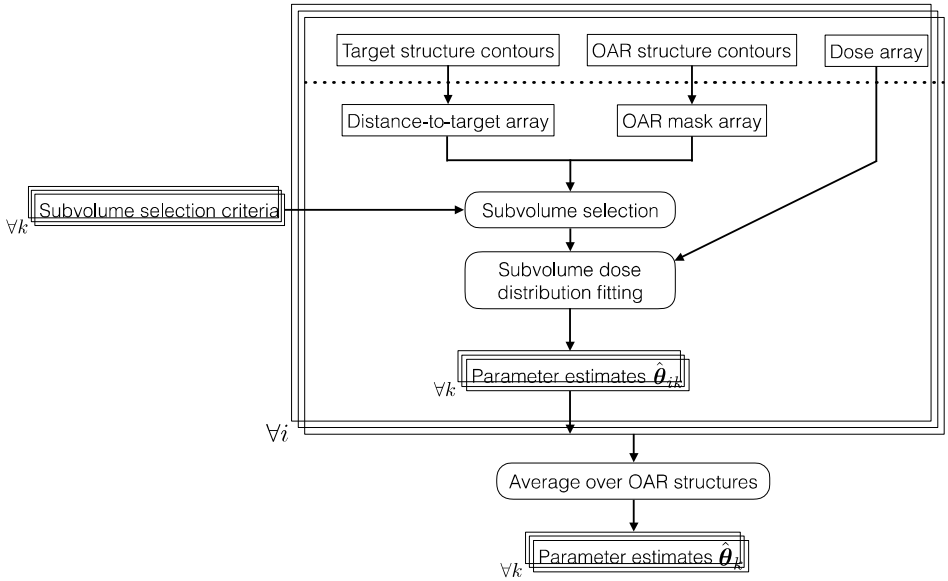

Figure 1. The training process for a single OAR type. The i and k indices label OAR structures (i.e. patients, symmetric pairs) and subvolumes, respectively.

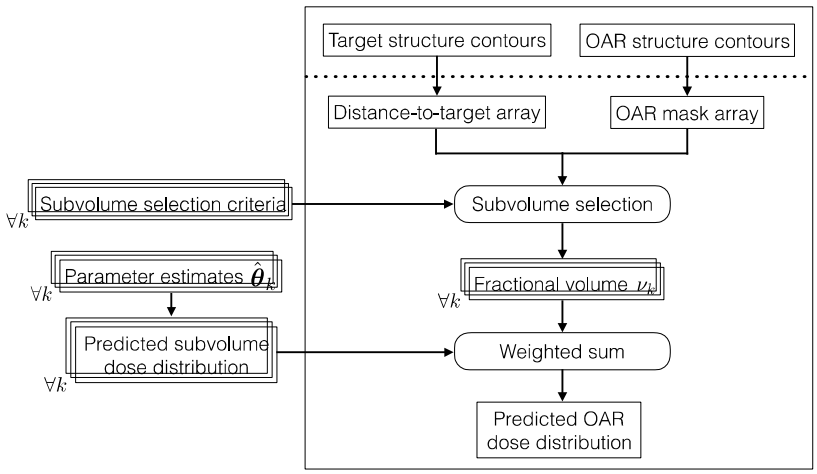

Figure 2. The prediction process for an OAR structure with greater than 100 voxels. The k index labels subvolumes.

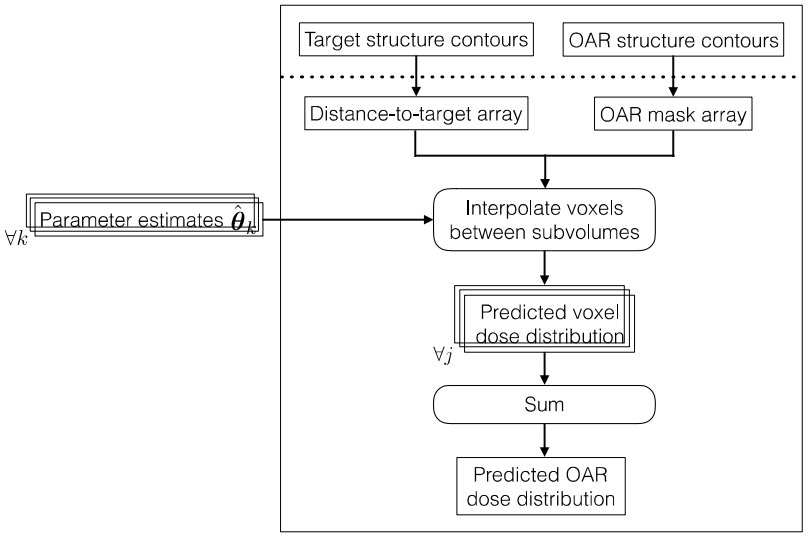

Figure 3. The prediction process for an OAR structure with fewer than 100 voxels. The k index labels subvolumes.