systematic and random setup errors [1]. This recipe does not explicitly account for interfraction time trends in tumor set-up, while such trends are observed for various tumor sites. In this work we propose 1) a novel characterization of set-up errors in a patient population with time trends, and 2) a margin recipe explicitly accounting for trends. The proposed formalism was evaluated for a large database of prostate cancer patients with time trends.

Material and Methods

The database contains daily set-up errors of 839 prostate cancer patients, measured in their 39 treatment fractions using implanted gold fiducials.

Errors in a patient population with time trends are described by normal distributions characterized by Σ, α and σ', with α the standard deviation of observed time trend slopes (mm/fraction) in the population and σ' describing the true random errors, i.e. errors relative to the patient's trend line. Figure 1 shows the set-up errors for a single database patient with a time trend. For the analyzed database, population parameters were: Σ = [2.5, 3.4, 3.5] mm, α = [-0.05, 0.07, 0.08] mm/fraction, and σ' = [1.9, 2.5, 2.6] mm for left-right, superior-inferior and anterior-posterior direction respectively.

Like in [1], the margin component for the random errors is given by $0.7\sigma'$. Similar to [1], we require for the margin component for the remaining errors (systematic and time trend errors) that 90% of the patients should be within the margin. The maximum set-up deviation during fractionated treatment, MD, of a patient with systematic set-up error, m, and time trend slope a is given by $MD=|m|+0.5(F-1)|a|$, with F the total number of fractions (Fig. 1). To establish the required margin, the MD distribution is first determined by random sampling from the m and a distributions (10^7 samples). The margin is then determined as the 90% cut-off point in the distribution.

For validation of the novel margin recipe and for comparison with van Herk's recipe we established for both recipes the percentage of patients outside the margin. That was done with the ellipsoid test where we looked for patients not fulfilling $\sum_i (MD_i/M_i)^2$ (with i denoting direction and M - margin size).
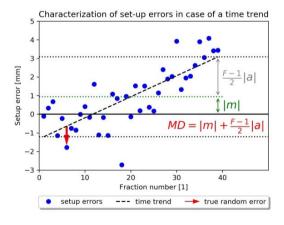

Figure 1. Maximum Deviation MD is derived from the middle position m and time trend slope a with the total number of fractions F.

Results

For the prostate database, margins calculated with van Herk's recipe were 1-2mm smaller than those established with the novel recipe (Table 1). The percentage of patients outside the novel margin was 9.8% (compared to 10% expected), while for van Herk's margin this was almost 26%.

	Van Herk Approach	Time Trend Approach
Left-Right margin [mm]	6.3 (7.6)	7.6 (8.9)
Superior-Inferior margin [mm]	8.5 (10.2)	10.9 (12.7)
Anterior-Posterior margin [mm]	8.8 (10.6)	11.0 (12.9)
Percentage of patients outside	25.9%	9.8%

Table 1. Comparison of margins for three directions for the clinical database. Margins in brackets include the random error component given as 0.7 σ'. Percentage of patients outside the margin is calculated for the systematic margin component.

Conclusion

Van Herk's CTV-PTV margin is not sufficient in case of time trends. We have proposed an extended recipe to fulfill the requirement that 90% of patients would indeed be irradiated with the prescribed dose when time trends are present. In case of no time trends, the modified recipe simplifies to van Herk's formula.

[1] van Herk et al., IJROBP 2000, Volume 47, Issue 4, Pages 1121–1135

PO-0996 A knowledge-based tool to estimate the gain of re-planning strategy for Head and Neck (HN) ART

E. Cagni[1], A. Botti[1], M. Orlandi[1], R. Sghedoni[1], E. Spezi[2], M. Iori[1]

[1]Azienda USL-IRCCS di Reggio Emilia, Medical Physics Unit, Reggio Emilia, Italy ; [2]Cardiff University, School of Engineering, Cardiff, United Kingdom

Purpose or Objective

This study aimed to investigate if a commercial knowledge-based tool for radiotherapy planning, RapidPlan (RP) (Varian Medical System, Palo Alto, CA), can be used to estimate the potential organ at risks (OARs) sparing in re-planning strategy for HN ART.

Material and Methods

A database of 45 HN VMAT plans, were used as training set for RP model. A second evaluation set, of 10 advanced oropharyngeal HN patients were randomly selected from the department database. All VMAT plans in the evaluation set were generated by means of RP module to treat 3 targets at dose levels of 69.96Gy/59.4Gy/54.12 Gy in 33 fractions using a SIB technique.

For each evaluation patient 2 CBCTs were extracted corresponded to 16th and the 26th fraction. In Velocity AI v.3.2 (Varian Medical System), the planning CT was registered with each CBCT using deformable registration algorithm, generating an Adaptive CT (ART-CT). For each ART-CT, the plan was re-calculated in Eclipse (delivered DVH) and RP predictions (RP DVH) were performed.

The gain of the re-planning was evaluated by comparing RP DVH with the delivered DVH for left and right parotid glands (PG), spinal cord and oral cavity. As a surrogate for the RP DVH we considered the line running in the middle of the predicted range. The restricted sum of the residuals (RSR) [Appenzoller et al. Med Phys. 2012] is used to measure the discrepancy between DVHs. To evaluate the feasibility of the method, the range of RP DVH estimations (RP uncertainties) were compared with the gain of the re-planning. The absolute sum of residual (ASR), considering both positive and negative difference in the sum, was used for this analysis.

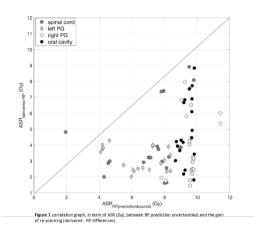Wilcoxon signed rank were used as statistical test.

Results

The RP model showed an average chi square of 1.06 ± 0.04 and coefficient of determination of 0.51 ± 0.11. Numerical values of RSR, that quantified the gain of re-planning, were reported in **Table 1**. The overall RSR (mean±1std), for all patient and both fractions, resulted 2.8±2.9Gy, 2.6±2.7Gy, 2.7±2.8Gy, 2.6±2.8Gy for spinal cord, left and right PG and oral cavity respectively. For 95% of the cases,

RP predicted a gain in the re-planning (RSR>0). No statistically difference resulted in RSR values between the 2 fractions (p=0.82).

	1	2	3	4	5	6	7	8	9	10
					fraction #16 RSR (Gy)					
					patient number					
spinal cord	1.1	2.5	0.0	7.1	8.7	3.6	2.5	1.4	4.4	3.9
left PG	2.0	4.4	0.5	1.6	4.4	0.6	0.1	0.0	0.0	0.4
right PG	4.6	5.5	0.0	1.2	2.7	0.4	0.1	0.9	0.0	6.7
oral cavity	0.3	7.5	0.0	0.0	3.5	0.1	0.1	0.0	0.0	5.5
					fraction #26 RSR (Gy)					
					patient number					
	1	2	3	4	5	6	7	8	9	10
spinal cord	0.9	1.1	0.0	7.2	8.1	3.1	2.2	1.3	0.0	3.6
left PG	4.0	2.6	0.5	3.2	1.7	0.7	2.0	0.0	0.7	4.6
right PG	5.8	1.1	0.1	1.7	1.0	0.9	1.8	0.6	0.0	7.8
oral cavity	1.3	5.7	0.0	0.4	3.3	0.1	0.2	0.0	0.0	5.3

Table 1 RSR values between delivered DVH and RP DVH that represent the gain of re-planning for spinal cord, left and right PG and oral cavity.

RP prediction uncertainties (RP bound ASR) resulted higher than the gain of re-planning (delivered- RP ASR), as is showed in **Figure1** by means of a correlation graph.. Statistical test confirmed this result (p<0.01)

Figure 1 correlation graph, in term of ASR (Gy), between RP prediction uncertainties and the gain of re-planning (delivered - RP differences).

Conclusion

In this study we have investigate the feasibility to use RP to estimate the potential gain of re-planning strategy for HN ART. Based on the analysis, DVHs predicted by RP can be used to estimate the potential OARs sparing when a new plan is performed. This information could be useful to assess the trigger point for a re-planning strategy. However, we found clinically relevant inaccuracies in RP predictions that limitate its application to HN ART. Therefore, further work is ongoing on RP model accuracy improvement.

PO-0997 A Synthetic Generative Adversarial Network for Semantic Lung Tumor Segmentation

V. Kearney[1], J.W. Chan[1], S. Haaf[1], S.S. Yom[1], T. Solberg[1]
[1]University of California UCSF, department of radiation oncology, San Francisco CA, USA
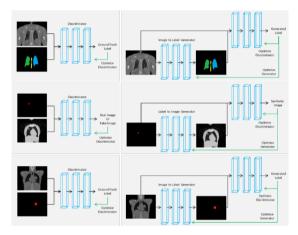
Purpose or Objective

To demonstrate the feasibility of a novel generative adversarial network (GAN) for synthetic abnormal pulmonary CT generation and semantic lung tumor segmentation.

Material and Methods

A 3D translational conditional GAN was implemented for synthetic image generation (label-to-CT) and segmentation (CT-to-label). Prior to synthetic image generation, a CT-to-label generator is given a CT image and trained to produce a binary mask of the left lung, right lung, heart, esophagus, spinal cord, and internal airways; a discriminator is trained to distinguish between "real" labels and synthetically generated "fake" labels. Once the network is conditioned, the label-to-CT synthetic image generator is trained by reversing the CT-to-label network and training the discriminator to perform the inverse task. The label-to-CT GAN is trained to generate arbitrary abnormal pulmonary CTs with various tumor characteristics, which are used for synthetic data augmentation. A final CT-to-label GAN is trained to generate binary tumor masks from a 4 to 1 mixture of synthetic and real pulmonary CTs for 200 epochs, and fine-tuned for 20 epochs on real pulmonary CTs. **Figure 1** shows the generator and discriminator components for all three GAN models. 208 stage I or stage II lung tumor patients previously treated with radiotherapy were used in this study. Patients with segmented hilar nodes were not included in this study. All algorithms were trained and hyperparameter tuned using 80% of the patients, and the remaining 20% were used to report final performance metrics. All models were distributed across two Nvidia V100 GPUs, and due to memory limitations, all images were resampled to 3x3x3 mm^3 and cropped to 128x128x64 voxels. To evaluate segmentation performance, all images were rescaled to their original dimensionality.

Figue 1. The preliminary GAN training workflow is shown for CT-to-label training of organs at risk (top). The label-to-CT workflow is shown which takes a binary tumor mask and generates arbitrary abnormal synthetic pulmonary CT variations (middle). The final CT-to-label GAN model is shown that generates realistic tumor masks given a CT image (bottom).

Results

The synthetic GAN model (synthetic-GAN) was compared to a GAN model (real-GAN) and V-Net model (real-VNet) using only traditional data augmentation (rotation, random cropping, elastic deformation, and translation). Among the 20 patients analyzed, the average dice scores and standard deviations were 0.82 ± 0.15, 0.71 ± 0.18, and 0.69 ± 0.16 for synthetic-GAN, real-GAN, and real-VNet respectively.

Conclusion

A synthetic conditional generative adversarial network was implemented that outperforms current state-of-the-art segmentation techniques for lung tumor segmentation. Furthermore, synthetically generated abnormal pulmonary images do not contain patient sensitive information and could be widely distributed to enhance cross institutional generalization.

PO-0998 Setup and range robustness recipes for skull-base meningioma IMPT using Polynomial Chaos Expansion

C. Ter Haar[1,2], S. Habraken[2,3], D. Lathouwers[1], R. Wiggenraad[3,4], S. Krol[3,5], Z. Perkó[1], M. Hoogeman[2,3]
[1]Delft University of Technology, Radiation Science & Technology, Delft, The Netherlands ; [2]Erasmus Medical Center Cancer Institute, Radiation Oncology, Rotterdam, The Netherlands ; [3]Holland Proton Therapy Center, Radiation Oncology, Delft, The Netherlands ; [4]Haaglanden Medical Center, Radiation Oncology, Leidschendam, The Netherlands ; [5]Leiden University