Use of knowledge based DVH predictions to enhance automated re-planning strategies in head and neck adaptive radiotherapy

Cagni Elisabetta[1,2], Botti Andrea[1], Chendi Agnese[1,3], Iori Mauro[1] and Spezi Emiliano[2]

[1]Medical Physics Unit, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy

[2]School of Engineering, Cardiff University, Cardiff, UK

[3]Department of Medical Physics, Alma Mater Studiorum Bologna University, Bologna, Italy

E-mail: elisabetta.cagni@ausl.re.it

Abstract

This study aimed to investigate if a commercial, knowledge-based tool for radiotherapy planning could be used to estimate the amount of sparing in organs at risk (OARs) in the re-planning strategy for adaptive radiotherapy (ART). Eighty head and neck (HN) VMAT Pareto plans from our institute's database were used to train a knowledge-based planning (KBP) model. An evaluation set of another 20 HN patients was randomly selected. For each patient in the evaluation set, the planning computed tomography (CT) and 2 sets of on-board cone-beam CT, corresponding to the middle and second half of the radiotherapy treatment course, were extracted. The original plan was re-calculated on a daily deformed CT (delivered DVH) and compared with the KBP DVH predictions and with the final KBP DVH after optimisation of the plan, which was performed on the same image sets. To evaluate the feasibility of this method, the range of KBP DVH uncertainties was compared with the gains obtained from re-planning. DVH differences and ROC curve analysis were used for this purpose. On average, final KBP uncertainties were smaller than the gain in re-planning. Statistical tests confirmed significant differences between the two groups. ROC analysis showed KBP performance in terms of area under the curve (AUC) values higher than 0.7, which confirmed a good accuracy in predicted values. Overall, for 48% of cases, KBP predicted a desirable outcome from re-planning, and the final dose confirmed an effective gain in 47% of cases. We have established a systematic workflow to identify effective OAR sparing in re-planning based on KBP predictions that can be implemented in an on-line, adaptive radiotherapy process.

Keywords: automatic planning, knowledge-based planning, multicriteria optimisation, adaptive radiotherapy

1. Introduction

External beam radiotherapy is gradually evolving towards real-time adaptive radiotherapy (ART) [1,2], which is becoming a new paradigm in radiation oncology [3]. ART has the clinical rationale of reducing normal tissue and organ at risk (OAR) toxicity and/or improving tumour control through plan adaptation [4,5]. The frequency of re-planning in patients with head and neck (HN) cancer was reported to vary from 32% to 70% depending on several criteria, such as weight loss, change in neck separation or poor immobilisation shell fit [6,7]. However, at present, there are still a number of technical limitations in applying ART as an automation and standardisation process [8]. As a consequence, ART is not widely utilised [7,8]. Indeed, in clinical practice, contouring and treatment planning processes are labour-intensive and use substantial resources. Some investigators are researching automated methods to predict the eventual need for re-planning, but more work is needed [9,10].

One of the on-going technical issues is the challenging decision of choosing the appropriate time in the process to re-plan a patient [11,12,13,14,15,16,17,18]. As anatomical deformation cannot be corrected by simple couch shifts these deformations should be managed by re-planning [8]. Generating a new plan with the same planning goals as the original plan within a clinically acceptable time and with minimal user intervention is another important technical challenge in ART. As the geometry of patient anatomy varies, OARs are shaped and positioned with respect to the targets and change from day to day. The gain in OAR sparing with a new plan is *a priori* unknown [11], and re-planning decisions are often based on the clinician's practical experience [11,13]. The "trigger point" is the time at which significant dosimetric variation for a specific parameter is present as an indicator for ART re-planning [14]. However, the re-planning process is expensive in resource terms as it requires a new computed tomography (CT) scan, new contours, and a new optimisation. For this reason, the trigger should be carefully chosen, balancing the time-consuming procedures with the gain from re-planning.

Knowledge-based planning (KBP) tools can generate estimated dose-volume histograms (DVHs) based on previous patient anatomy and dose distributions [19,20]. The KBP methods are generally equivalent to expert planners in terms of plan quality but preliminary results indicate that they are significantly more efficient timewise [19]. Currently, KBP is frequently used in clinical practice to drive new IMRT planning based on a database of prior clinical plan data and other sources of knowledge, such as treatment trade-off and clinician experience. Studies published in recent years have demonstrated that KBP allowed a general improvement in (a) inter-patient consistency of the treatment plans, (b) intrinsic quality and (c) efficiency (time and workflow) of the planning process applied to different anatomical sites using manually generated treatment plans [21,22,23,24,25,26], including the head and neck [27,28].

Previously published studies have demonstrated that the efficacy of KBP was influenced by the quality of the data used for the training process, the regression applied to build the predictive models and the consistency between new cases and the population used for the training [20]. KBP training generally involves using a database of manually generated treatment plans that can suffer from plan quality variation and inconsistency [24]. To maximise the performance of KBPs, Pareto plan solutions can be used to train the model [29,31]. In the studies published to date [21,22,23,24,25,26,27,28], KBP was employed to assist the planner to achieve optimal dose distributions when new plans were created.

The clinical implementation and use of KBP optimization models is a rapidly changing subject and previous studies have applied KBP for help in replanning for adaptive radiation therapy [32]. However, to the best of our knowledge, none of the published studies have investigated if the accuracy of a KBP prediction is suitable for OAR sparing in an ART application. In this work, we investigate the possibility of using the KBP method as part of the ART process to estimate the potential gain given by OAR sparing during the treatment course for HN cases. We hypothesise that such relationships,

if significant, could be used to establish the need for plan adaptation based on OAR sparing and to automate the process of re-planning itself.

2. Methods and Materials

2.1 Clinical Data

In line with samples used in previous work [28,30], a dataset of 100 Head and Neck (HN) VMAT patients previously treated at AUSL-RCCS of Reggio Emilia was selected for our study. Treatment plans following two different fractionation schemes were included in our work, 69.96 Gy/59.4 Gy/54.12 Gy in 33 fractions (46 patients) and 66 Gy/60 Gy/54 Gy in 30 fractions (54 patients), both schemes using a simultaneous integrated boost (SIB) technique. For all plans, the goal was to deliver 100% of the prescribed dose to 95% of every PTV. Each plan was generated using a previously configured KBP model for HN patients trained on manually produced clinical plans. This model is discussed below (Section 2.4). All plans were generated with the Eclipse Treatment Planning System (TPS) (Varian Medical Systems, Palo Alto, CA) using 3 fully coplanar arcs with collimator rotated to 30°, 315° and 90°, with 6 MV energy. For each treatment fraction, an on-board cone beam CT (CBCT) scan was acquired before treatment delivery to assess patient setup and anatomical variation. All CBCT images were automatically saved in an ARIA database (Varian Medical Systems, Palo Alto, CA).

2.2. KB RapidPlan planning tool

RapidPlan is a knowledge-based automatic planning (KBP) solution integrated in the Eclipse TPS. For each new patient, RapidPlan predicts the most likely dose-volume histograms (DVHs) to occur based on the specific patient's anatomy in terms of structure set geometries. Predicted DVHs are then used to establish dose-volume objectives and weights for automated plan optimisation. DVH prediction in RapidPlan is based on a statistical model that is generated from the principal component analysis of anatomic and dosimetric features obtained from plans of previously treated patients. Therefore, the quality of RapidPlan DVH predictions depends on the quality of the plans used to train the model. Extensive descriptions of model configuration are provided in the Varian reference manual [33] and previous publications [21,22,23,24,25,26,27,28].

2.3 Multicriteria Optimisation (MCO) Trade-off module

The Multicriteria Optimisation (MCO) approach is based on trade-off exploration modules and is implemented in the Eclipse TPS. In MCO, a range of different Pareto solution plans is generated based on a selection of optimisation

objectives. The priority of each objective may vary from plan to plan but all plans belong to a 'Pareto surface', which means that is not possible to improve an objective without compromising another objective in the trade-off. The user can explore the trade-offs along the Pareto surface and select the plan that best fulfils the treatment goals. With the use of slider bars, dynamic DVHs and dynamic 3D dose distributions, the TPS allows users to visually review and evaluate plans along the Pareto surface in real-time [31,33].

2.4 MCO-KBP model

The 100 manual plans in our database were first randomly divided into two groups: KBP training (80 plans) and KBP evaluation (20 plans). Using the training set, KBP DVH prediction models were created for the following OARs: brainstem, spinal cord, left parotid gland, right parotid gland, mandible, oral cavity, oesophagus, and larynx. The KBP model was built following the guidelines provided by the manufacturer [33]. **Figure 1a** outlines the workflow of the KBP model generation stage.

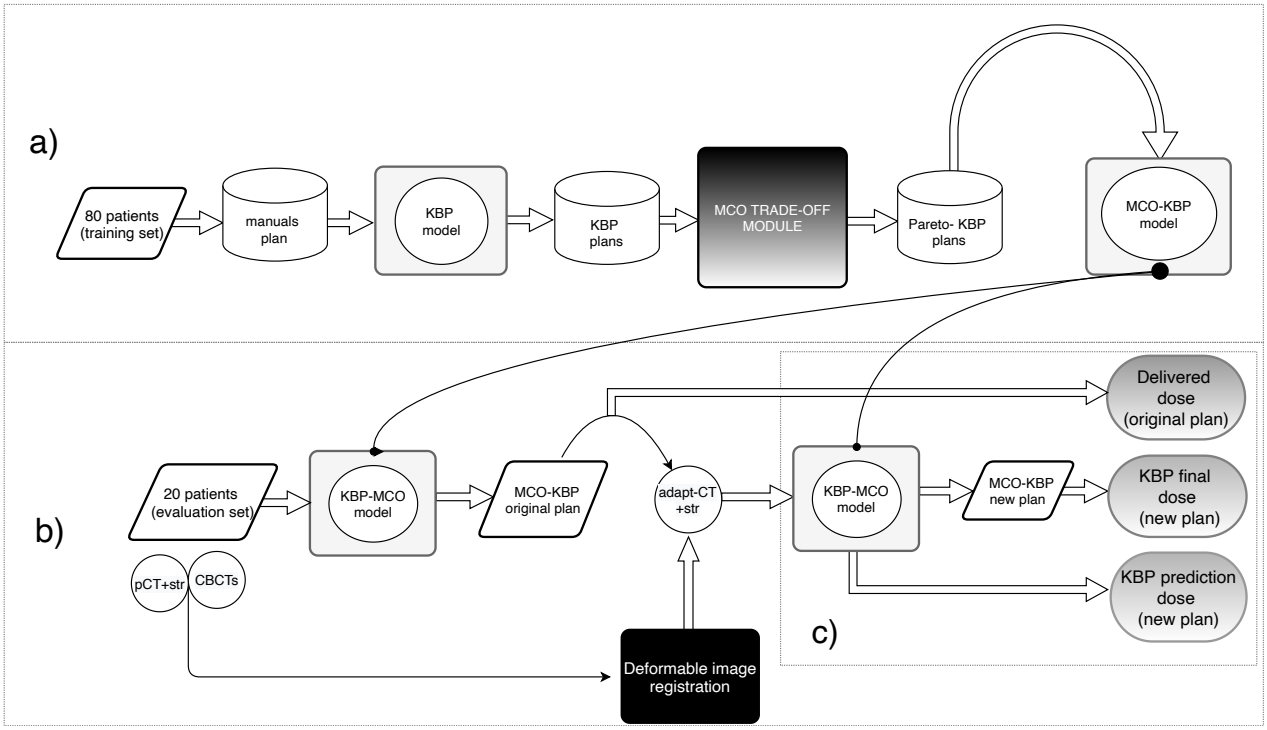

Figure 1. The workflow of the study was divided into 3 parts**:** MCO-KBP models creation a) and application of the gain of the re-planning estimation using the KBP tool b) and c). In detail: 100 HN patients were selected for the study. a) 80 patients were chosen to train the KBP model. To improve the quality of the model, the training set was automatically re-planned to use the KBP model and refined using the MCO tool based on a single wish list. The final MCO-KBP plans generated in this process were then used to train the model. b) 20 patients were considered in the evaluation set for the experiment. For each patient's planning CT (pCT), the structure set and CBCT of fractions 16 and 26 were considered. Using a deformable image registration tool, adapt-CT, deformed structures for both fractions were generated. c) An automated plan was created using the KBP_MCO model on pCT. The plan was recalculated using the adapt-CT of each fraction generating the delivered dose. The KBP_MCO model was used to generate a KBP prediction and final dose from the adapt-CT of each fraction if a new plan was created.

All 80 patients used in the training set were automatically re-optimised using the KBP model. The solution found after optimisation without manual intervention was used as a starting plan in the MCO module. A wish list of objectives to fulfil was used to consistently select a solution on the Pareto surface in the MCO module. This wish list was created together with radiation oncologists at our institution, based on previous clinical experience and planning protocols [35,36,37,38]. In the MCO module, once the goal of an objective in the wish list was obtained, the optimiser fixed the achieved value (slider restrictor) as a constraint and continued with the next objective.

Once the MCO module had completed the optimisation for every objective on the wish list, the DVHs and the spatial distribution of the dose that resulted from the wish list trade-off process were selected from the Pareto surface, and the final dose was calculated. Due to the use of the same optimisation scheme for all patients, plan generation was highly consistent across the entire cohort, with no plans adjusted by the planning team.

All plans generated with the KBP and MCO combination were used to train a new, highly consistent MCO-KBP model (MCO-KBP) for the previously selected structures. The optimisation constraints for this model were unchanged with respect to the KBP model.

2.5 Delivered DVHs for the evaluation set

In the evaluation dataset, the MCO-KBP model was used for each patient to create an automatically generated plan following the same fractionation scheme as the original treatment. This procedure is outlined in **Figure 1b.** Two CBCT scans acquired before each treatment fraction were extracted for each patient, corresponding to the 16^{th} fraction (around half-way through the treatment course) and the 26^{th} fraction (about three quarters of the way through the treatment course). For each selected fraction, an adapted CT image [39] was generated in Velocity AI v.4.0 (Varian Medical Systems, Palo Alto, CA) through deformable registration of the planning CT (pCT) on the selected CBCT. The same HU values as the pCT were used, to avoid bias in dose calculations. The transformation used was a cubic B-spline, the interpolator was computed with a bi-linear function, and the optimiser was based on the steepest gradient descent [39]. For the selected set of patients, no artifacts were evident in the CT image. No image pre-processing was done before the deformable image registration. The OAR structures were automatically propagated from the planning CT to the adapted CT. The accuracy of registration was verified by means of a visual inspection of deformed CT and structure sets. This was combined with a visual check of deformable vector field (DVF) that was performed in Velocity AI after each registration.

The adapted CT images with modified structure sets were then imported into the Eclipse TPS, and the plan created with the MCO-KBP model was recalculated for the adapted CT. The dose and DVH of the plan for this adapted CT will be referred to as the 'delivered DVH' (d_DVH).

2.6 MCO-KBP DVH model prediction validation

For each patient in the evaluation dataset, KBP predictions were performed on both adapted CTs using the MCO-KBP model. This section of the workflow is outlined in **Figure 1c**. For each organ, RapidPlan presents two predicted DVH lines representing the ±1 standard deviation DVH confidence limits. For the validation group, the mean of these two predicted DVH curves, pKBP_DVH, was compared with the corresponding DVH achieved after the optimisation, fKBP_DVH [40]. Following international guidelines [35,36,37], specific DVH endpoints for validation were considered: maximum dose (Dmax) for spinal cord, brainstem and mandible and mean dose (Dmean) for parotids, oral cavity, larynx and oesophagus, as used in our clinical practice.

2.7 Gain from re-planning and KBP prediction uncertainty

From the Varian Manual [33], the normal prediction bounds shown by the KBP tool indicated 68% probability (±1standard deviation (SD)) that the final DVH should fall between the bounds. Based on this assumption, we can consider the predicted KBP uncertainty as 1 SD, the distance between the pKBP_DVH (section 2.6) and one of the prediction bounds. Thus, for each OAR *j* and case *i*:

$$\text{pKBP uncertainty}_{ji} = ASR\left(upper\ bound\ KBP\_DVH_{ji} - pKBP\_DVH_{ji}\right) \quad (1)$$

ASR was the absolute sum of residuals (ASR) and was used to quantify the distance between DVHs:

$$\text{ASR} = \left(\sum_{D=0}^{\infty} |DVH_1(D) - DVH_2(D)| \cdot \Delta D\right) \quad (2)$$

where $\text{DVH}_{1j}(D)$ *and* $\text{DVH}_{2j}(D)$ refer to the DVHs for which the distance was quantified.

After the KBP prediction was obtained, the optimisation was performed and the dose calculated without manual intervention, generating a final KBP DVH (fKBP_DVH).

In line with **eq. 1** and **eq. 2**, the final KBP uncertainty was defined as the difference between predicted and final DVH:

$$\text{fKBP uncertainty}_{ij} = ASR\left(fKBP\_DVH_{ij} - pKBP\_DVH_{ij}\right) \quad (3)$$

where for each OAR *j* and case *i*, fKBP_DVH refers to the final KBP DVH after optimisation, with ASR and pKBP_DVH as defined above.

The gain from re-planning was calculated in term of sum of residuals (SR), between delivered DVH (section 2.5), $\text{d\_DVH}_{ij}(D)$, and predicted KBP, $\text{pKBP\_DVH}_{ij}(D)$ or final DVH after optimisation, $\text{fKBP\_DVH}_{ij}(D)$, when the new plan is created:

$$SR_{ij} = \left(\sum_{D=0}^{\infty}\left(\text{d\_DVH}_{ij}(D) - \text{KBP\_DVH}_{ij}(D)\right)\cdot \Delta D\right) \quad (4)$$

for each OAR *j* and case *i*. We referred respectively to predicted gain values (pSRs) or to final gain values (fSRs), when pKBP_DVH or fKBP_DVH was considered in **eq. 4**.

The approach to use sum of residuals to estimate the difference of DVHs is similar to the one proposed by Appenzoller et al. [41]. However, in the original study, only positive differences between DVHs were considered for detection of suboptimal plans (restricted sum of residuals). We accounted for both positive and negative differences, as both are important to detect improvement/detriment in the re-planning phase. To automate the gain from re-planning, predicted gain values (pSR) were plotted against the final optimisation gain (fSR).

To evaluate the feasibility of our method, the KBP uncertainty (**eq. 1** and **eq. 3**) for each OAR and case was compared against the gain from re-planning. To use the same metric in the comparison, ASR, was used to quantify the distance between DVHs in the gain of replanning, using **eq.2**. Thus, for each OAR *j* and each case *i*:

$$ASR_{ij} = \left(\sum_{D=0}^{\infty}\left|d\_DVH_{ij}(D) - KBP\_DVH_{ij}(D)\right|\cdot \Delta D\right) \quad (5)$$

We referred respectively to ASR predicted gain values (pASRs) or to final gain values (fASRs), when pKBP_DVH or fKBP_DVH was considered in **eq. 5**.

2.8 Discriminant analysis

The receiver operating characteristic (ROC) analysis curve (referred in this manuscript as "discriminant analysis") was used to quantify predicted KBP performance measures in this experiment. With these analyses, four domains were highlighted:

1. TP: positive pSR values predict significant OAR sparing, and re-planning demonstrates improved OAR DVH with positive fSR re-planning values.

2. True negatives (TN): negative or null pSR values predict no OAR sparing, and re-planning demonstrates no improvements in OAR DVH with negative or null fSR re-planning values.

3. FP: positive pSR values predict significant OAR sparing, but re-planning demonstrates no improvements in OAR DVH with negative or null fSR re-planning values.

4. False negatives (FN): negative pSR values predict no OAR sparing, but re-planning demonstrates improved OAR DVH with positive fSR re-planning values.

The ROC curve was used to choose the best predicted operating point (OP) that gives the best trade-off between the sensitivity, or TP rate, and specificity, or 1- FP rate, of KBP predictions [42].

2.9 Statistical Analysis

Wilcoxon two-sided signed rank tests were used to assess the statistical significance of the observed differences between KBP predictions and final endpoints and SRs, to test the difference in gain between the two fractions and to test differences between ASR values and KBP uncertainty values. The differences were considered significant when $p<0.05$.

3. Results

3.1 Quality of the MCO-KBP model

The quality of the generated MCO-KBP model was evaluated by checking the model's goodness-of-fit statistics for each structure such as the coefficient of determination (R^2 (between 0 and 1: the larger, the better)) and the average Pearson's chi-square (χ^2 (the closer to 1, the better)), as suggested by RapidPlan guidelines [33]. These parameters, together with the number of potential outliers (also known as influential points), are reported in **Table 1** for all models. No particular trends were observed for χ^2 and R^2. A mean χ^2 of 1.11±0.05 and a mean R^2 of 0.83±0.10 were found.

Table 1. Goodness of the prediction models in terms of coefficient of determination, R^2, average Pearson's chi square, χ^2, and number of potential outliers (model analytics, MA, suggested plans to be removed and plans to be checked)

Structure	R^2	χ^2	# Outliers (MA)
Brainstem	0.84	1.05	3
Oesophagus	0.83	1.06	0
Larynx	0.82	1.18	3
Mandible	0.83	1.08	5
Oral Cavity	0.87	1.07	2
Parotids	0.82	1.05	0
Spinal Cord	0.54	1.10	0

The potential outliers identified the Varian Model Analytic (MA) tool [34] were evaluated on a case-by-case basis. Since all plans inserted in the model were Pareto plans, these cases were regarded as not actually outliers, and it was verified that there were no anomalous anatomical differences compared to the rest of the population in the model to cause such categorisation. The plans identified as potential outliers by MA, were representative of an "uncommon" patient anatomy with respect to the training set population, but these plans remained clinically suitable as they were created using the MCO module.

The differences between predicted and final KBP endpoints for the OARs considered are reported in **Table 2**. The differences were calculated for every predicted and final endpoint of each organ and patient, then the mean and SD values over the patient population were derived. Mean difference ± 1SD between the pKBP and fKBP endpoints were 1.8 ± 4.5 Gy, 1.2 ± 5.3 Gy. 0.8 ± 1.1 Gy, 0.6 ± 0.9 Gy, -1.3 ± 1.8 Gy. 1.4 ± 2.4 Gy, 0.9 ± 2.1 Gy and -3.4 ± 4.7 Gy for spinal cord, brainstem, right and left parotid, oral cavity, oesophagus, larynx and mandibula, respectively. Wilcoxon signed rank tests confirmed the differences across the range of endpoints (pKBP and fKBP) were not statistically significant ($p > 0.05$).

Table 2 Endpoint differences (Dmean and Dmax) between prediction (pKBP) and final DVH (fKBP) for all 20 patients (2 CBCTs for patient) of the evaluation set in term of mean value±standard deviation.

	Spinal cord	**Brainstem**	**Right parotid**	**Left parotid**	**Oral cavity**	**Oesophagus**	**Larynx**	**Mandibula**
	Dmax [Gy]	Dmax [Gy]	Dmean [Gy]	Dmean [Gy]	Dmean [Gy]	Dmean [Gy]	Dmean [Gy]	Dmax [Gy]
pKBP_DVH	28.5±4.7	28.6±9.0	21.5±8.6	23.0±6.4	34.3±10.7	16.5±7.1	32.0±5.2	64.9±6.2
fKBP_DVH	26.7±4.6	27.3±8.5	20.7±9.0	22.4±6.6	35.5±11.7	15.0±6.3	31.1±6.2	68.2±8.1
difference	1.8±4.5	1.2±5.3	0.8±1.1	0.6±0.9	-1.3±1.8	1.4±2.4	0.9±2.1	3.4±4.7

3.2 Gain from re-planning

Figure 2 shows a comparison of the DVHs across an individual patient case (patient #7 fraction 16), considered as a patient example. The predicted pKBP_DVH, the bounds of the predicted pKBP_DVH (1 SD of the KBP), and the final fKBP_DVH and the d_DVH are represented for the following organs: spinal cord, brainstem, left parotid, mandible, larynx, oesophagus, oral cavity and right parotid.

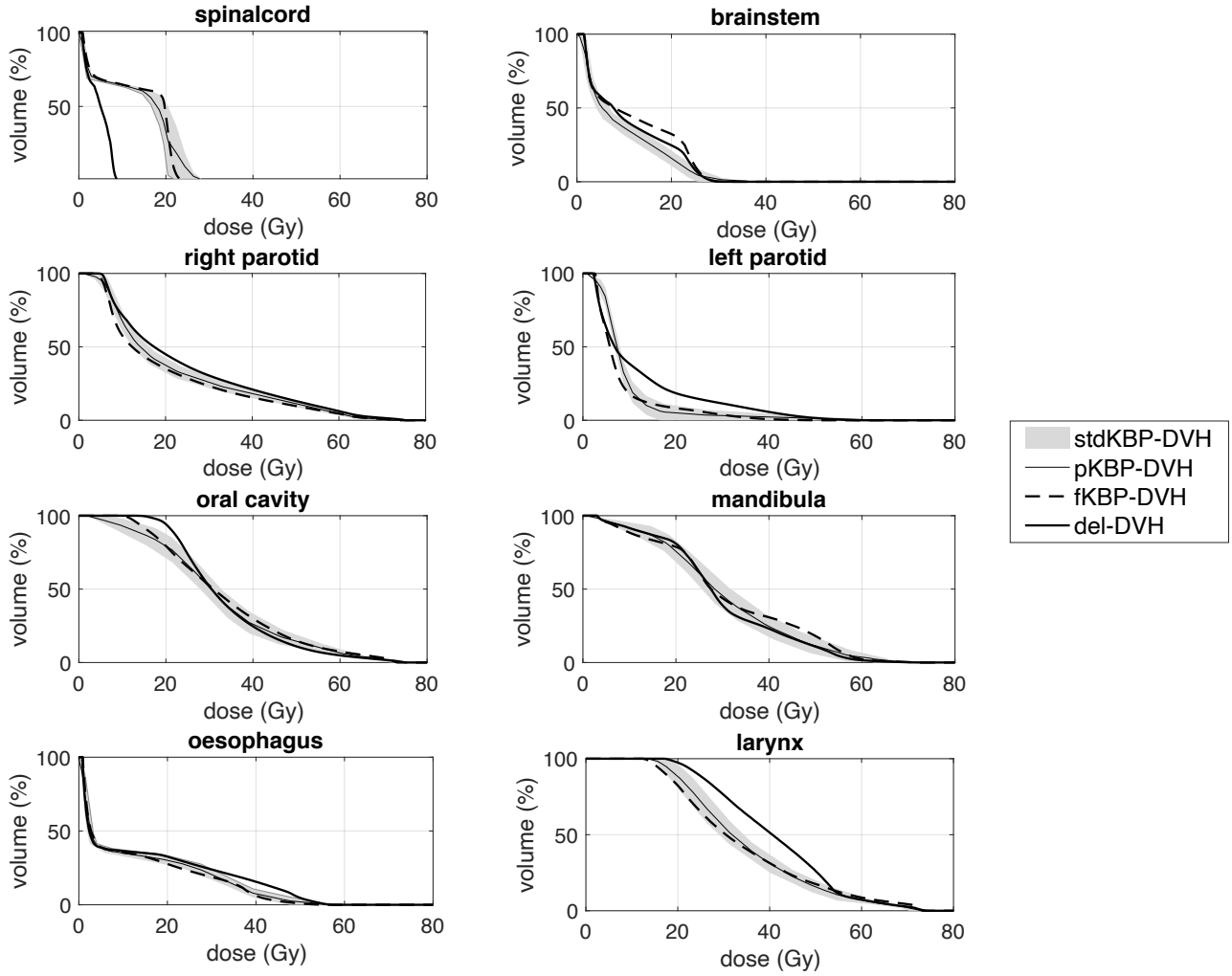

Figure 2 KBP-predicted DVH bounds (stdKBP), mid KBP-predicted DVH (pKBP_DVH), final KBP DVH (fKBP_DVH) and delivered DVH (d_DVH) for each organ for patient #7 fraction 16.

The overall fSR and pSR (mean±1SD) for all patient OARs and the fractions were 0.07±2.73and 0.08±2.98, respectively. A significant difference was found between pSR and effective fSR (p=0.03). For each OAR, the mean and standard deviation of fSR and pSR values (the latter in brackets) were -1.34 ±2.15 (-0.97 ±2.30), 0.10 ±1.13 (0.07±1.20), 1.54 ±4.04 (0.70±3.81), 0.59 ±2.96 (0.03 ±2.96), -0.20 ±2.72 (1.07±2.61), -0.55 ±1.47 (1.48 ±3.00), 0.59±2.17 (-0.85±3.13) and -0.27 ±3.08 (-1.18 ±3.04) for spinal cord, brainstem, left parotid, right parotid, mandible, oesophagus, oral cavity and larynx, respectively. The fSR and pSR values were similar between the two fractions. The Wilcoxon two-sided signed rank test confirmed significant differences for fSR values of the 16^{th} and 26^{th} fractions at $p<0.01$. No significant difference was observed in the pSR distribution of the two fractions.

For 48% of the cases (N=310), KBP predicted a positive gain from re-planning (pSR>0). Final DVHs confirmed the effective gain fSR>0 in a similar percentage, 47%, of cases. **Figure 3** shows a boxplot of the comparison between pSRs

and fSRs for each organ for all 40 cases. Overall, there was no observed trend between median fSRs and median pSRs. For brainstem, parotids, oesophagus, and larynx, median fSRs were higher than their respective pSRs, as reported in **Figure3**.

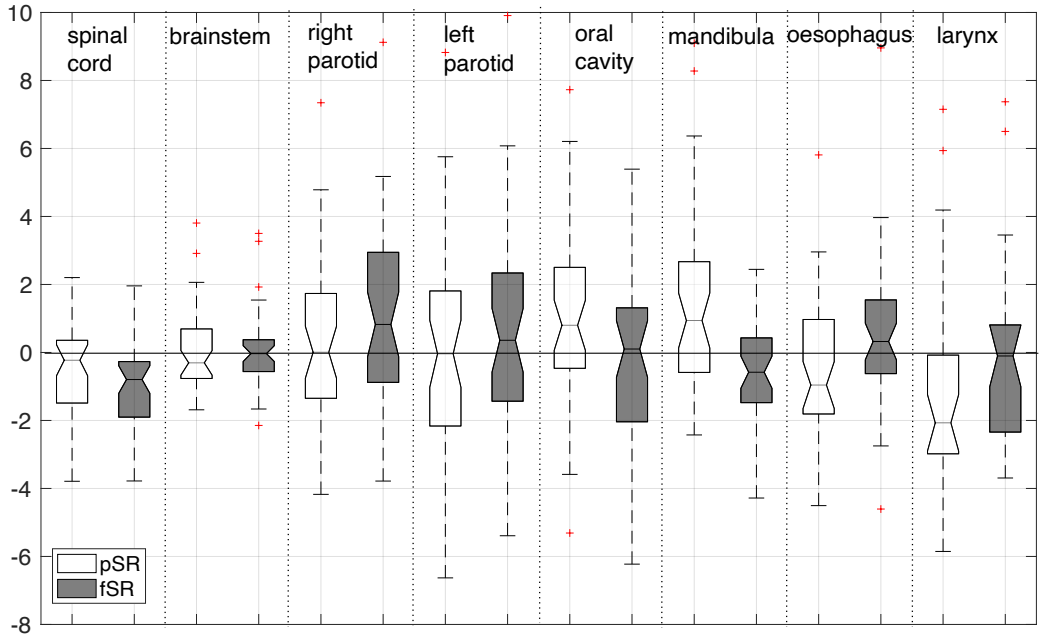

Figure 3 Boxplot of predicted and final gain in terms of pSRs (white) and fSRs (shaded) for each organ for all 20 delivered adapted_CBCT and KBP plans. The central mark represents the median value, whereas the edges of the boxes are the 25th and 75th percentiles, respectively. The whiskers extend to the maximum distance, and the crosses represent individual outliers.

3.3 KBP model uncertainties

Ideally KBP prediction uncertainties should be significantly smaller than, or at least comparable to the predicted gain. However, even if the model was trained with Pareto optimal plans, a relevant KBP prediction uncertainty is still present. For each OAR and each case, we compared the pKBP uncertanty values versus the pASR, the latter representing the predicted gain of replanning quantified using ASR (Section 2.7). These values are plotted in the correlation graph reported in **Figure 4a**. Each marker represents a plan of the evaluation set for which the pASR was higher (circle) or lower (cross) than the pKBP uncertainty. Overall, the mean pASR values were similar to KBP uncertainties and the pASR SD slightly higher than KBP uncertainties, with values of 2.78±2.13 and 2.38±0.79 respectively. A Wilcoxon two-sided signed rank statistical test confirmed that there was not a significant difference between the two groups (p=0.56). From **Figure 4**, it is possible to observe that the correlation between KBP uncertainties and pASR was strictly related to OAR and a single item. The ideal situation in **Figure 4a** would be to see all points well above the diagonal line of unity, indicating a pASR value significantly larger than the KBP uncertainty (for each case). However, many items in **Figure 4a** fall below the

diagonal. This finding means that in those cases KBP uncertainties were higher than predicted gains in terms of ASR. It was found that 15/40, 20/40, 19/40, 20/40, 20/40, 15/40, 23/40 and 15/30 points for pASR were higher than the corresponding pKBP uncertainty, for spinal cord, brainstem, left and right parotid, oral cavity, mandible, oesophagus, and larynx respectively. This is represented in **Figure 4** as a square in the lower right corner. When the gain was lower/higher than uncertainties for more than half of the items the square is filled with black/gray colour; in the case of equality the square was white.

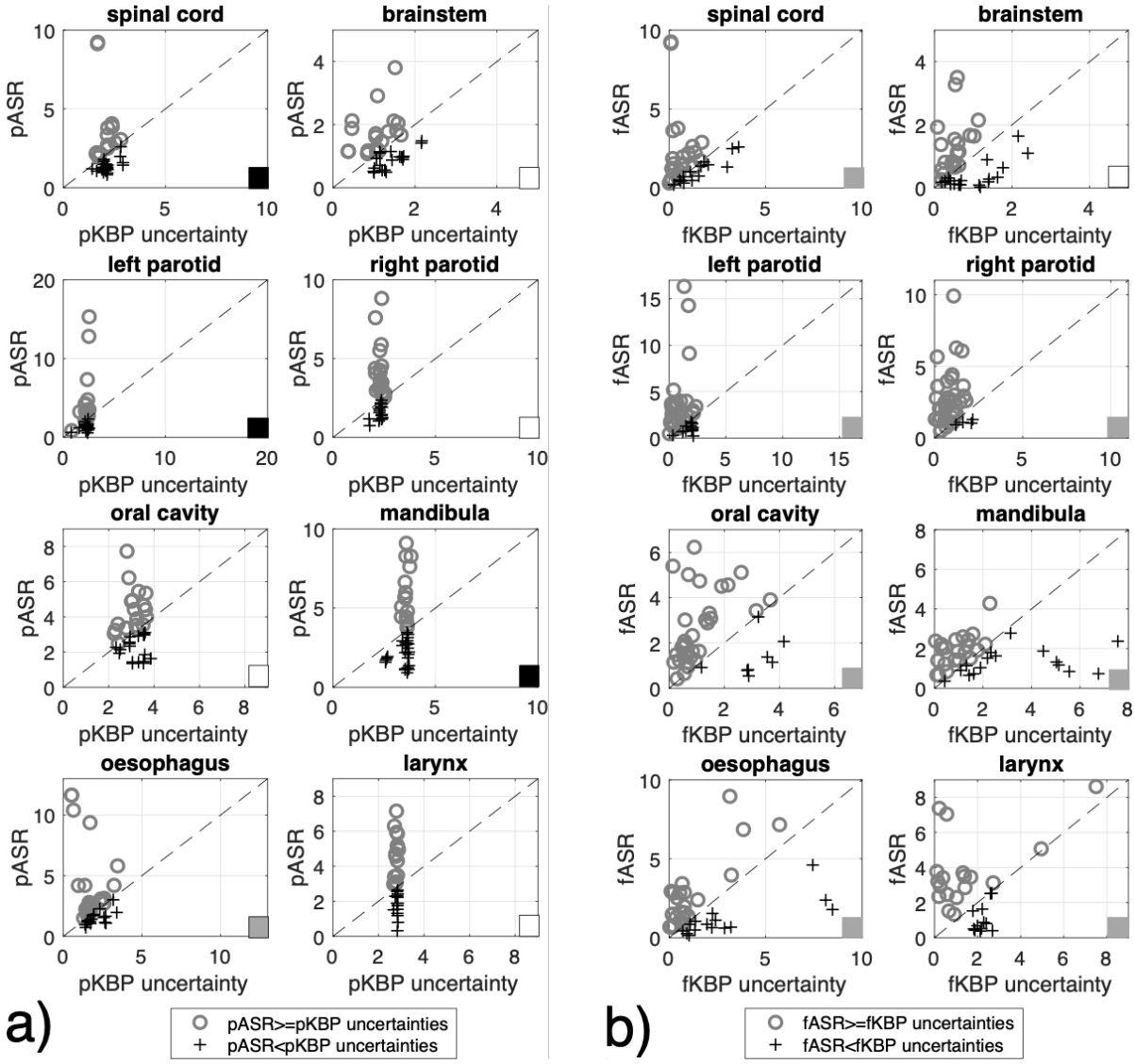

Figure 4 a) Correlation graphs of 1 SD predictions of pKBP uncertainties versus pASR from re-planning, for each OAR model. **b)** Correlation graphs of fKBP uncertainties versus fASR from re-planning, for each OAR model. Each marker represents a plan for the evaluation set for which the gain is higher (circle) or lower (cross) than the KBP uncertainty. This is represented in each panel of the figure as a square in the lower right corner. When the gain resulted lower/higher that uncertainties for more than half of the items the square is filled black/gray colour; in the case of equality the square was filled in white.

In **Figure 4b** the correlation between fKBP uncertainties and final gains, quantified as fASR, is shown. In comparison with **Figure 4a,** where pKBP uncertainties were considered, fewer points fall below the bisector. This means that the fKBP uncertainties were smaller than final gain, in terms of ASR. Overall, mean and SD values of 2.12 ±2.02 and 1.40 ±1.61 were obtained for fASR and fKBP uncertainties respectively, with p-values $<<0.01$. This confirms that the difference between the two distributions is statistically significant. In this case we found that 26/40, 20/40, 28/40, 36/40, 31/40, 22/40, 25/40 and 18/30 cases for predicted gain were higher than the corresponding KBP uncertainties, for spinal cord, brainstem, left and right parotid, oral cavity, mandible, oesophagus, and larynx respectively.

3.4 Discriminant analysis.

Figure 5 shows the correlation graph between pSR and fSR for the 20 patients and 2 fractions considered (N=40 cases). Overall, the two groups were deemed to be correlated with a coefficient of 0.72 ($p<0.01$). Single OAR correlation coefficients and p-values were 0.90 ($p<0.01$), 0.91 ($p<0.01$), 0.57 ($p=0.18$), 0.76 ($p=0.04$), 0.71 ($p=0.04$), 0.83 ($p=0.01$), 0.72 ($p=0.04$) and 0.42 ($p=0.30$) for the spinal cord, brainstem, parotids (left and right), oral cavity, mandible, oesophagus, and larynx, respectively.

From discriminant analysis, the OP values for each of the OARs were 0.20, 0.09, -0.61, 0.70, 0.78, 1.67, 0.03 and -2.15. These OPs represent guidelines for clinical decision making using pSR of KBP values to obtain a real gain after plan optimisation (fSR>0). **Figure 5** reports, for each OAR, the pSR and fSR. The OPs are represented by a vertical line in each graph for each organ. It is possible to observe that, for various OARs, many fSRs are positive for pSR values greater than OPs (marked as circles).

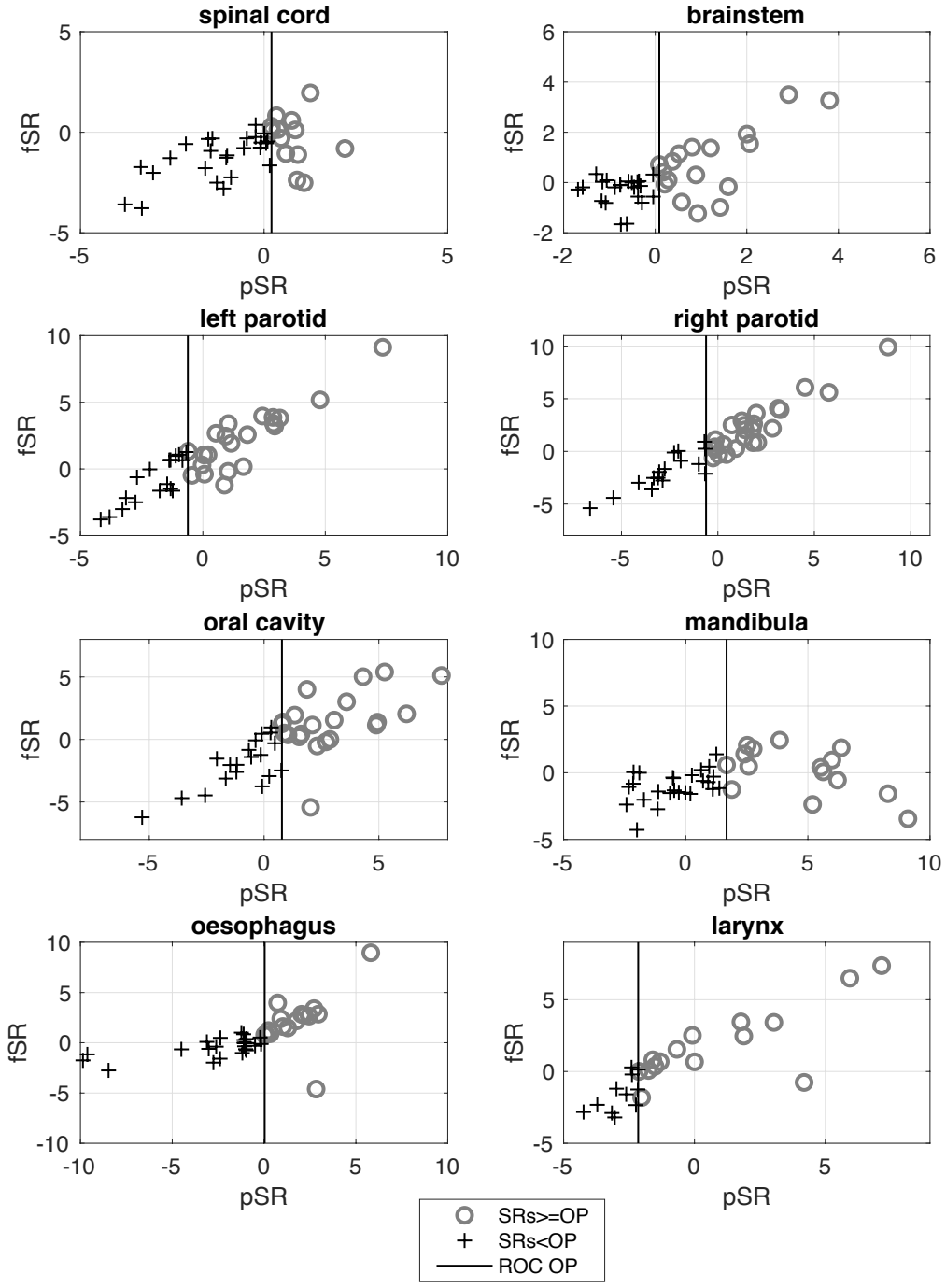

Figure 5 Correlation graph of predicted KBP gain pSR and fSR values for the 20 patients, where both fractions are considered (40 adapted CTs). OPs found from the ROC analysis are represented by a vertical line in each graph for each organ. Each marker represents a plan for the evaluation set for which the pSR were higher/right (circle) or lower/left (cross) of the OP value (vertical line).

In **Figure 6**, ROC curves are shown separately for each OAR. The corresponding areas under the curve (AUCs) give an effective measure of the accuracy of the pSR prediction and were 0.824, 0.704, 0.974, 0.863, 0.772, 0.821 and 0.886 for the spinal cord, brainstem, parotids, oral cavity, mandible, oesophagus, and larynx, respectively.

Overall, among the 310 cases (in 10 cases the larynx was missing), TP were 140, TN were 104, FP were 36 and FN were 30. From the ROC analysis, the sensitivity and specificity of the overall model were 0.69 and 0.78, respectively. The associated accuracy was 0.74, and the estimation error was 0.26.

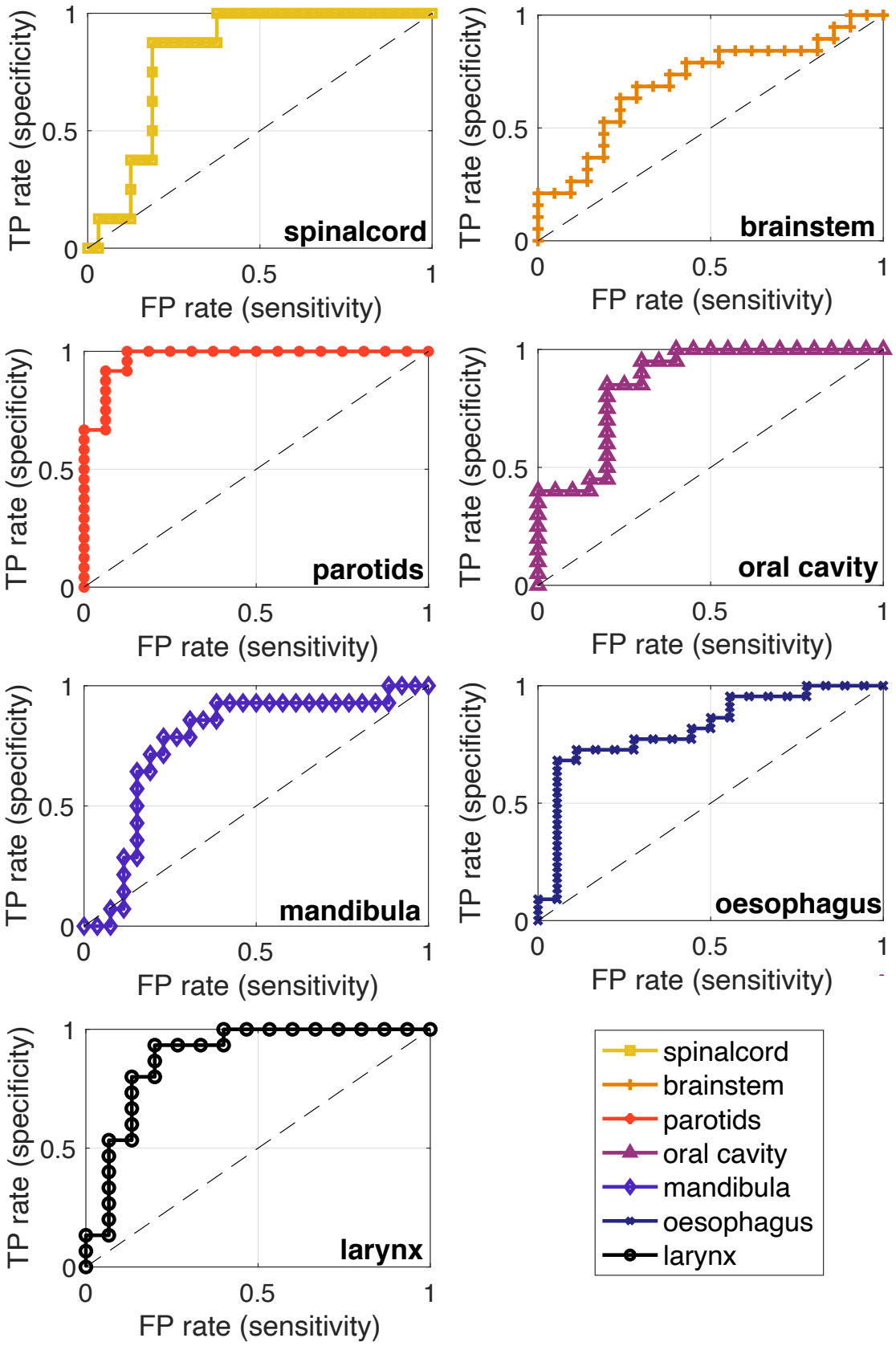

Figure 6 ROC curve for each OAR. The corresponding AUCs were 0.824, 0.704, 0.974, 0.863, 0.772, 0.821 and 0.886 for the spinal cord, brainstem, parotids, oral cavity, mandible, oesophagus, and larynx, respectively.

4. Discussion

In this study, we have investigated the use of a KBP tool for ART and evaluated the predicted gain from re-planning for OARs. The KBP model was trained using Pareto-optimal treatment plans that were all automatically generated using the Eclipse MCO module with a uniquely prioritised, objective optimisation list, avoiding the use of manually generated plans to train the model as these generally suffer from variation in quality and inconsistencies among different planners. To assess the inherent accuracy of RapidPlan predictions for ART re-planning, the gain from re-planning was quantified by comparing DVHs of the original plan recalculated on patient anatomy, imaged at two different time points during treatment, with predicted and final KBP DVH obtained with a new plan on the same image set.

Discriminant analysis performed in this study allowed an estimation of the KBP predictive power for the effective gain from re-planning, with an AUC value greater than 0.7 (**Figure 6**) for all OARs, confirming that mid-line DVH could be used as a good surrogate for prediction values as previously suggested in other works [28,40]. ROC curve analysis established the best cut-off for predicted values for clinical purposes (OPs (Section 2.8)). These OPs were found to be positive (range: [0.03, 1.67]) for 5 OAR models, except for the parotids (-0.61) and larynx (-2.15).

We used adapted-CTs based on deformable image registration between planning CT and CBCT respectively, for fraction 16 and fraction 26. The validation of image registration algorithm for clinical use remains a challenging task in ART, as their accuracy depends on the complexity and quality of the images used in the registration task [43,44]. In the absence of standardized tools, the accuracy of registration in our study was verified by means of a visual inspection. We carefully visually assessed deformed images (adapted-CTs), deformed structures on adapted-CTs and magnitude of the deformable vector fields. Since the investigations were performed directly on adapted-CTs by comparing predicted and final KBP DVH with the original plan recalculated on adapted-CT, our analysis was not directly influenced by the registration uncertainties.

An optimal prediction model needs a prediction range (predicted uncertainty) that is as small as possible. Said range is defined in RapidPlan as 1 SD from the predicted result. A larger prediction range means that the model has larger uncertainties. In the MCO-KBP model training, we used consistently generated Pareto-optimal plans to build models. However, a significant statistical deviation from predicted DVH bounds is still present as reported in **Figure 2** for an example patient from the evaluation set. Since we avoided the use of manual plans to build the model, the deviation is not related to the quality of the suboptimal plans but only to the intrinsic limitations of the model.

The KBP prediction error bounds led to significantly higher pKBP uncertainty, as reported in **Figure 4a** compared with the predicted gain in term of ASR for several cases. Both quantities (pASR and KBP uncertainties) were dependent on the OAR KBP model (**Figure 4**). Overall, pASR of pKBP uncertainties and the ASR of predicted gains when a new plan

was used were very close in average values and Wilcoxon signed rank tests confirmed no significant differences between the two groups ($p>0.05$). When considering the final values after optimisation this relationship improved, showing fASR values higher than corresponding fKBP uncertainty values for the majority of cases (**Figure 4b**).

Overall, predicted, and final gains, pSR and fSR, were correlated ($p<0.01$) and showed a similar proportion of cases with positive gain predicted (48%) with respect to positive actual gain (47%).

In the process of building the KBP prediction model, outliers are defined as training plans that could result in undesirable bias in the models. In the absence of general rules, some criteria to better identify and manage these possible outliers have been previously published [28, 29,34]. In this study, all plans used to build the model were Pareto-optimal with consistent trade-offs between all treatment objectives. Therefore, we expected not to find any dosimetric outliers in our training set. However, the Varian MA tool still identified some instances of outliers in our training dataset (**Table 1**). Those potential outliers could have been found due to the limited range of geometric information in the input data and/or intrinsic limits of the models, as reported in a previous study by Cagni et al. [29].

Varian's RapidPlan guide indicates a minimum number of 20 plans to train a model. However, Boutilier et al [30] have shown that for prostate cancer DVH prediction, RapidPlan needs at least 75 plans to achieve good prediction accuracy. Fogliata et al. [28] considered 83 patients for building a KBP model with RapidPlan and the models were validated on 20 HN patients. In line with this approach, our model was built on 80 cases for training and 20 cases for validation.

The model quality was evaluated by checking the goodness of fit statistics for each structure, with the coefficient of determination R^2, with observed values higher than 0.8 for all structures except spinal cord, and the average Pearson's chi square χ^2, with all observed values less than 1.1 (**Table 1**). These results are in line with the KBP models of goodness of fit reported in other studies [20,27,28]. Moreover, differences in terms of model predicted and final endpoints (Dmax and Dmean) were not statistically significant (**Table 2**), confirming the goodness of model prediction.

Sum of residuals (SR) between entire DVHs were used as the metric to quantify predicted and final OAR gain. From a radiobiological perspective, SR gives interesting information about global DVH and mean dose variation, generally considered for parallel organs, such as the parotids, larynx, oral cavity, or oesophagus [45,46]. On the other hand, for serial organs such as the spinal cord, only the high dose region of the DVH is correlated with radiation-induced complications [45,46]. In this work, the spinal cord had the worst (negative) gain among all OARs with a mean final SR of -1.34, but it did not provide information about the radiobiological integrity of the organ since there was no indication of how its maximum delivered dose changed. For this reason, we also considered specific DVH endpoints to summarize the RapidPlan prediction performance, as described in **Table 2.**

5. Conclusions

We have demonstrated the feasibility of using knowledge-based tools to establish the need for plan adaptation based on OAR sparing to help automate the process of re-planning. This study has shown that the prediction uncertainties of knowledge-based planning tools trained with Pareto-optimal plans are sufficiently low for such tools to be used in adaptive radiotherapy. The approach described in this work has the potential to be implemented in an on-line adaptive radiotherapy process.

Acknowledgments

The authors would like to thank Prof Geraint Lewis, Ms Iona Foster, and Ms Valeria Trojani for useful comments on the manuscript.

Conflict of interest

The authors have no conflicts to disclose.

References

1. Keall P, Poulsen P, Booth JT 2019 See, Think, and Act: Real-Time Adaptive Radiotherapy. *Semin Radiat Oncol* 29(3) 228-235.
2. Green OL, Henke LE, Hugo GD 2019 Practical Clinical Workflows for Online and Offline Adaptive Radiation Therapy *Semin Radiat Oncol* 29(3) 219-227.
3. Jaffray DA 2012 Image-guided radiotherapy: from current concept to future perspectives *Nat Rev Clin Oncol* 9 688-699.
4. Langen KM, Jones DT 2001Organ motion and its management *Int J Radiat Oncol Biol Phys* 50 265-278.
5. Yan D, Vicini F, Wong J, Martinez A 1997 Adaptive radiation therapy *Phys Med Biol* 42(1) 123-32.
6. Heukelom J, Fuller CD 2019 Head and Neck Cancer Adaptive Radiation Therapy (ART): Conceptual Considerations for the Informed Clinician *Semin Radiat Oncol* 29(3) 258-273
7. Ahunbay EE, Peng C, Godley A, Schultz C, Li XA 2009 An on-line replanning method for head and neck adaptive radiotherapy *Med Phys* 36(10) 4776–4790.
8. Liu C, Kim J, Kumarasiri A, Mayyas E, Brown SL, Wen N, Siddiqui F, Chetty IJ 2018 An automated dose tracking system for adaptive radiation therapy *Computer Methods and Programs in Biomedicine* 154 1-8.
9. Gensheimer MF, Quynh-Thu Le 2018 Adaptive radiotherapy for head and neck cancer: Are we ready to put it into routine clinical practice? *Oral Oncology* 86 19-24.

10. Guidi G, Maffei N, Meduri B et al 2016 A machine learning tool for re-planning and adaptive RT: a multicenter cohort investigation *Phys Med* 32(12) 1659-1666.
11. Zhang P, Simon A, Rigaud B, Castelli J, Ospina Arango JD, Nassef M, Henry O, Zhu J, Haigron P, Li B, Shu H, De Crevoisier R 2016 Optimal adaptive IMRT strategy to spare the parotid glands in oropharyngeal cancer *Radiother Oncol* 120(1) 41-7.
12. Stoll M, Giske K, Debus J, Bendl R, Stoiber EM 2014 The frequency of replanning and its variability dependent on the modification of the replanning criteria and IGRT correction strategy in head and neck IMRT *Radiat Oncol* 9 175.
13. Brouwer CL, Steenbakkers RJ, Langendijk JA, Sijtsema NM 2015 Identifying patients who may benefit from adaptive radiotherapy: Does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help? *Radiother Oncol* 115(3) 285-294.
14. Huang H, Lu H, Feng G, et al. 2015 Determining appropriate timing of adaptive radiation therapy for nasopharyngeal carcinoma during intensity-modulated radiation therapy *Radiat Oncol* 10 192.
15. Ahn PH, Chen CC, Ahn AI, et al 2011 Adaptive planning in intensity- modulated radiation therapy for head and neck cancers: single-institution experience and clinical implications *Int J Radiat Oncol Biol Phys.* 80 677-685.
16. Lee C, Langen KM, Lu W, et al 2008 Assessment of parotid gland dose changes during head and neck cancer radiotherapy using daily mega-voltage computed tomography and deformable image registration Int J Radiat Oncol Biol Phys 71 1563-1571.
17. Stoiber EM, Schwarz M, Huber PE, Debus J, Bendl R, Giske K 2013 Comparison of two IGRT correction strategies in postoperative head_and_neck IMRT patients *Acta Oncol* 52(1):183–186.
18. Graff P, Kirby N, Weinberg V, Chen J, Yom SS, Lambert L, Pouliot J 2013 The residual setup errors of different IGRT alignment procedures for head and neck IMRT and the resulting dosimetric impact *Int J Radiat Oncol Biol Phys* 86(1) 170–176.
19. Ge Y, Wu QJ 2019 Knowledge-based planning for intensity-modulated radiation therapy: A review of data-driven approaches *Med Phys* 46(6) 2760-2775.
20. Fogliata A, Cozzi L, Reggiori G, Stravato A, Lobefalo F, Franzese C, Franceschini D, Tomatis S, Scorsetti M 2019 RapidPlan knowledge-based planning: iterative learning process and model ability to steer planning strategies Radiat *Oncol* 14(1) 187.
21. Fogliata A, Wang PM, Belosi F, Clivio A, Nicolini G, Vanetti E, Cozzi L. 2014 Assessment of a model-based optimisation engine for volumetric modulated arc therapy for patients with advanced hepatocellular cancer *Radiat Oncol* 9:236.
22. Hussein M, South CP, Barry MA, Adams EJ, Jordan TJ, Stewart AJ, Nisbet A 2016 Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy. Radiother Oncol 120 473–9.

23. Fogliata A, Nicolini G, Clivio A, Vanetti E, Laksar S, Tozzi A, Scorsetti M, Cozzi L 2015 A broad scope knowledge-based model for optimisation of VMAT in esophageal cancer: validation and assessment of plan quality among different treatment centers *Radiat Oncol* 10 234.
24. Fogliata A, Nicolini G, Bourgier C, Clivio A, De Rose F, Fenoglietto P, et al 2015 Performance of a knowledge-based model for optimisation of volumetric modulated arc therapy plans for single and bilateral breast irradiation *PLoS One* 10(12) e0145137.
25. Snyder KC, Kim J, Reding A, Fraser C, Gordon J, Ajlouni M, et al 2016 Development and evaluation of a clinical model for lung cancer patients using stereotactic body radiotherapy (SBRT) within a knowledge-based algorithm for treatment planning *J Appl Clin Med Phys* 17(6) 263–75.
26. Foy JJ, Marsh R, Ten Haken RK, Younge KC, Schipper M, Sun Y, et al 2017 An analysis of knowledge-based planning for stereotactic body radiation therapy of the spine Pract Radiat Oncol 7(5) e355–60.
27. Tol JP, Delaney AR, Dahele M, Slotman JB, Verbakel WFAR 2015 Evaluation of a knowledge-based planning solution for head and neck cancer *Int J Radiat Oncol Biol Phys* 91(3) 612–20.
28. Fogliata A, Reggiori G, Stravato A, Lobefalo F, Franzese C, Franceschini D, Tomatis S, Mancosu P, Scorsetti M, Cozzi L. 2017 RapidPlan head and neck model: the objectives and possible clinical benefit *Radiat Oncol* 12(1):73.
29. Cagni E, Botti A, Wang Y, Iori M, Petit SF, Heijmen BJM 2018 Pareto-optimal plans as ground truth for validation of a commercial system for knowledge-based DVH- prediction Phys Medica 55 98–106.
30. Boutilier JJ. 2016 Sample size requirements for knowledge-based treatment planning. Med. Phys. 43 (3).
31. Miguel-Chumacero E, Currie G, Johnston A, Currie S 2018 Effectiveness of Multi-Criteria Optimisation-based Trade-Off exploration in combination with RapidPlan for head & neck radiotherapy planning *Radiat Oncol* 13(1) 229.
32. Castriconi R, Fiorino C, Passoni P, Broggi S, Di Muzio N, Cattaneo GM , Calandrino R. 2020 Knowledge-based automatic optimization of adaptive early-regression-guided VMAT for rectal cancer. Physica Medica 70:58-64.
33. Varian Medical Systems, Inc. 2017 Eclipse photon and Electron algorithms reference guide Palo Alto CA 221–31 www.myvarian.com
34. Varian Model Analytics 2018 https://modelanalytics.varian.com/
35. Loco-regionally Advanced Oropharyngeal Carcinoma: Radiation Therapy Oncology Group (RTOG) Trial 0022 - Phase I/II Study of Conformal and Intensity Modulated Irradiation for Oropharyngeal Cancer. Activation Date: February 1, 2001
36. Loco-regionally Advanced Nasopharyngeal Carcinoma: Radiation Therapy Oncology Group (RTOG) Trial 0225- A Phase II Study of Intensity Modulated Radiation Therapy (IMRT) +/ Chemotherapy for Nasopharyngeal Cancer February 21, 2003.

37. Loco-regionally Advanced Nasopharyngeal Carcinoma: Radiation Therapy Oncology Group (RTOG) Trial 0615 - A phase II study of concurrent chemoradiotherapy using three-dimensional conformal radiotherapy (3D_CRT) or intensity modulated radiation therapy (IMRT) +bevacizumab (BV) for locally or regionally advanced nasopharyngeal cancer. Activation Date: December 13, 2006.
38. Lee NY, Ferris RL, Beck T, Harrington K, Haddad RI, Bourhis J, Tahara M, Geraldes M, Dimitry Nuyten DSA, Goldberg Z, Cohen EEW 2017 JAVELIN head and neck 100: A phase 3 trial of avelumab in combination with chemoradiotherapy (CRT) vs CRT for 1st-line treatment of locally advanced squamous cell carcinoma of the head and neck (LA SCCHN) *Journal of Clinical Oncology* 35 15.
39. Velocity AI v.4.0 instruction for use 2018 Varian Medical System www.myvarian.com
40. Tol JP, Dahele M, Delaney AR, Slotman BJ, Verbakel WF 2015 Can knowledge-based DVH predictions be used for automated, individualized quality assurance of radiotherapy treatment plans? *Radiat Oncol* 10(1) 234.
41. Appenzoller LM, Michalski JM, Thorstad WL, Mutic S, Moore KL 2012 Predicting dose-volume histograms for organs-at-risk in IMRT planning *Med Phys* 39(12) 7446-61.
42. Bradley AP 1997 The use of the area under the ROC curve in the evaluation of machine learning algorithms Pattern Recognition 30(7) 1145-1159.
43. Brock KK. Image Processing in Radiation Therapy 1st ed. CRC Press; 2013.
44. Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. Med Phys. 2017;44(7): e43-e76.
45. Bentzen M 2010 Quantitative analyses of normal tissue effects in the clinic (QUANTEC): An introduction to the scientific issues Int J Radiat Oncol Biol Phys 76(3) S3–S9.
46. Beltran M, Ramos M, Rovira JJ, Perez-Hoyos S, Sancho M, Puertas E, Benavente S, Ginjaume M, Giralt J. 2012 Dose variations in tumor volumes and organs at risk during IMRT for head-and-neck cancer. Journal of applied clinical medical physics 13 (6), 101-111, 2012.