Delft University of Technology

Accurate assessment of a Dutch practical robustness evaluation protocol in clinical PT with pencil beam scanning for neurological tumors

Rojo-Santiago, Jesús; Habraken, Steven J.M.; Lathouwers, Danny; Méndez Romero, Alejandra; Perkó, Zoltán; Hoogeman, Mischa S.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Original Article

Accurate assessment of a Dutch practical robustness evaluation protocol in clinical PT with pencil beam scanning for neurological tumors

Jesús Rojo-Santiago [a,b,*], Steven J.M. Habraken [a,b], Danny Lathouwers [c], Alejandra Méndez Romero [a,d], Zoltán Perkó [c], Mischa S. Hoogeman [a,b]

[a] Erasmus MC Cancer Institute, University Medical Center Rotterdam, Department of Radiotherapy, The Netherlands; [b] Department of Medical Physics & Informatics, HollandPTC, Delft, The Netherlands; [c] Delft University of Technology, Department of Radiation Science and Technology, Delft, The Netherlands; [d] Department of Radiation Oncology, HollandPTC, Delft, The Netherlands

ABSTRACT

Background and purpose: Scenario-based robust optimization and evaluation are commonly used in proton therapy (PT) with pencil beam scanning (PBS) to ensure adequate dose to the clinical target volume (CTV). However, a statistically accurate assessment of the clinical application of this approach is lacking. In this study, we assess target dose in a clinical cohort of neuro-oncological patients, planned according to the DUPROTON robustness evaluation consensus, using polynomial chaos expansion (PCE).

Materials and methods: A cohort of the first 27 neuro-oncological patients treated at HollandPTC was used, including realistic error distributions derived from geometrical and stopping-power prediction (SPP) errors. After validating the model, PCE-based robustness evaluations were performed by simulating 100.000 complete fractionated treatments per patient to obtain accurate statistics on clinically relevant dosimetric parameters and population-dose histograms.

Results: Treatment plans that were robust according to clinical protocol and treatment plans in which robustness was sacrificed are easily identified. For robust treatment plans on average, a CTV dose of 3 percentage points (p.p.) more than prescribed was realized (range +2.7 p.p. to +3.5 p.p.) for 98% of the sampled fractionated treatments. For the entire patient cohort on average, a CTV dose of 0.1 p.p. less than prescribed was achieved (range −2.4 p.p. to +0.5 p.p.). For the 6 treatment plans in which robustness was clinically sacrificed, normalized CTV doses of 0.98, 0.94(7)[1], 0.94, 0.91, 0.90 and 0.89 were realized. The first of these was clinically borderline non-robust.

Conclusion: The clinical robustness evaluation protocol is safe in terms of CTV dose as all plans that fulfilled the clinical robustness criteria were also robust in the PCE evaluation. Moreover, for plans that were non-robust in the PCE-based evaluation, CTV dose was also lower than prescribed in the clinical evaluation.

© 2021 The Authors. Published by Elsevier B.V. Radiotherapy and Oncology 163 (2021) 121–127 This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

The fundamental assumption underlying the planning target volume (PTV) concept in radiotherapy is the static dose cloud approximation [1], i.e., the invariance of dose under small shifts. However, it does not apply to proton therapy (PT) with pencil beam scanning (PBS) and cannot be extended to include proton stopping-power prediction (SPP, range) uncertainty either [2,3]. With the wide clinical introduction of PT with PBS in recent years, scenario-based mini-max robust optimization [4–7] is increasingly used to ensure adequate dose of the clinical target volume (CTV) in the presence of geometrical uncertainties, e.g., patient setup and alignment errors, and SPP errors. In mini-max robustness optimization, robustness settings might still be based on conventional CTV-PTV margin recipes [8,9], although the issue of finding optimal settings has been addressed [10–12].

Since the PTV is also used to evaluate the CTV dose in conventional radiotherapy, scenario-based robust optimization requires novel robustness evaluation strategies [13]. Various approaches have been proposed [2,14–16]. The HollandPTC clinical robustness evaluation protocol is based on consensus within the Dutch Proton Therapy (DUPROTON) group [17]: CTV dose is prescribed to the voxel-wise minimum (VWmin) dose [18] based on 28 evaluation scenarios and taking into account different combinations of geometrical and range errors in line with the clinically used robust optimization settings. In reference [17], this approach is validated

* Corresponding author at: Doctor Molewaterplein 40, 3015GD Rotterdam, The Netherlands.
E-mail address: j.rojosantiago@erasmusmc.nl (J. Rojo-Santiag.

[1] To indicate that the value is just below the constraint of 0.95, it was not rounded to two decimal digits, but it is displayed with the third decimal digit between parentheses.

and compared to a PTV-based approach in conventional radiotherapy. It has, however, two limitations: (i) the error scenarios are fixed and limited in their number while the actual errors follow continuous distributions and (ii) it does not address systematic and random geometrical errors separately, which are different for each treatment and for each fraction. A priori, it remains uncertain how the clinical robustness settings relate to the actual patient errors, and it is not guaranteed that this approach leads to adequate robustness in treatment planning [19–21]. Furthermore, it is unclear what is the impact of a VWmin CTV underdose with the current clinical protocol.

The aim of this study is to assess the performance of the clinical robustness evaluation protocol in a clinical cohort of neuro patients, using comprehensive patient- and treatment plan-specific modelling of the dose. This cohort includes not only treatment plans that meet clinical robustness evaluation constraints (clinically robust plans) but also plans in which robustness had to be compromised to meet the organs-at-risk (OAR) constraints (clinically non-robust plans).

Realistic distributions for the systematic and random geometrical errors are derived from quality assurance (QA) and patient data, while the distribution of proton range errors is based on literature [22–24]. Our method involves polynomial chaos expansion (PCE), which is first validated for realistic error distributions and then applied to simulate 100.000 complete fractionated treatments for each patient, thus providing an accurate calculation of the delivered dose; the probability of achieving adequate CTV dose, both at the level of an individual patient and the patient population; as well as comprehensive probabilistic metrics of the current clinical robustness evaluation protocol. Hence, this population-based analysis allows to improve current clinical treatment planning and can be extended to future neuro-oncological treatments.

Methods and materials

Patient data

The first 27 patients treated at HollandPTC for meningioma, grade-I glioma, grade II-III oligodendroglioma with 1p/19q co-deletion and grade-II astrocytoma with isocitrate dehydrogenase (IDH) mutation [25] between September 2018 and September 2019 were included. One patient was treated for two targets in one plan and was therefore excluded. Prescribed doses (D_{pres}) were 45 Gy(RBE) (1 case), 50.4 Gy(RBE) (15 cases), 54 Gy(RBE) (2 cases), 59.4 Gy(RBE) (7 cases), in 1.8 Gy(RBE) fractions and 60 Gy(RBE) (1 case) in 2 Gy(RBE) fractions, using a constant relative biological effectiveness (RBE) of 1.1. Dose was prescribed to the VWmin near-minimum CTV dose, i.e., VWmin- $D_{98\%,CTV} \geq 95\% D_{pres}$. A clinical constraint was also put on the VWmax near-maximum CTV dose [17]: VWmax-$D_{2\%,CTV} \leq 107\% D_{pres}$, in line with widely used ICRU recommendations [26,27].

T1 (Gadolinium-enhanced), T2 and FLAIR sequences of magnetic resonance imaging (MRI) were acquired for each patient. A rigid registration of MRI and a single-energy planning computed tomography (SECT) scan, with $0.68 \times 0.68 \times 1$ mm CT resolution,

in head first supine position was used to facilitate gross tumor volume (GTV) definition and OARs delineation. The median GTV was 101.5 cc (range 14.0–364.3 cc). OARs, i.e., brainstem, spinal cord, optic chiasm, optic nerves, cochleas, retinae, lenses, pituitary, lacrimal glands and the hippocampi were delineated in line with the European Particle Therapy Network (EPTN) delineation atlas for neuro-oncology [28,29].

Treatment planning

Clinical PT treatment plans in RayStation (version 7, RaySearch Labs, Sweden) treatment planning software (TPS) with patient-specific non-coplanar arrangements of two or three beams were available for all patients. 7 patients with Titatinum surgical clips were planned with a single-field uniform dose (SFUD) approach, while multi-field optimization (MFO) was used for all other cases. An isotropic 3 mm setup robustness setting was used, i.e., the vector length of the isocenter shifts of the optimization scenarios was 3 mm, based on 98% population coverage robustness recipes [30]. Prior to the first patient treatment (with no clinical patient setup data available yet), systematic and random geometrical errors of 1 mm each, which are slightly larger than the actual clinical setup errors (cf., section 2.3), were assumed. For the relative SPP error, a range robustness setting of 3% was used in line with literature [11,12].

Clinical robustness evaluation was based on 28 combined range and geometrical error scenarios [17] in line with the clinically used robust optimization settings (3 mm setup error and 3% range). The VWmin-$D_{98\%,CTV}$ constraint was met in 20 patients with an average VWmin-$D_{98\%,CTV}/D_{pres}$ value of 0.96 (range 0.95–0.97). However, robust target dose was sacrificed in 6 cases to spare critical OARs, in particular the brainstem and the optic nerves, specifically in the one planned to 60 Gy(RBE), one planned to 50.4 Gy(RBE) and in 4 of the 7 planned to 59.4 Gy(RBE). The normalized VWmin-$D_{98\%,CTV}/D_{pres}$ values for these patients (patient 21 to 26 in Fig. 3A) were $0.94(9)$[1], 0.92, 0.90, 0.89, 0.88 and 0.87 respectively. The VWmax-$D_{2\%,CTV}$ constraint was met in all cases. Relevant patient characteristics, dosimetric parameters and planning constraints are listed in Table 1 and Table 2.

Treatment uncertainties

Geometrical errors are primarily due to (i) variations in patient setup and anatomy, and (ii) registration and isocenter misalignment errors. In a fractionated treatment, they can be split into a systematic and a random component, the former of which is fixed during the complete treatment while the latter is different for each treatment fraction [8].

Patient setup data was available for all 26 patients included in the analysis. It was obtained from residual setup errors in pairs of orthogonal planar kilovolt (kV) images, acquired in each fraction after the last 6D correction of the robotic treatment couch, just prior to the treatment delivery. Systematic components were obtained as the standard deviation (1 SD) of the residual mean setup errors, while the random errors are defined by the root-

Table 1
Patient characteristics per subgroup, showing the median and the range (between brackets) for the CTV volume, the $D_{2\%,CTV}$ and $D_{98\%,CTV}$ in the scenario with no uncertainties (nominal) and from the clinical robustness evaluation.

Patient group	CTV (cc)	$D_{2\%,CTV}$ Gy(RBE)	VWmax-$D_{2\%,CTV}$ Gy(RBE)	$D_{98\%,CTV}$ Gy(RBE)	VWmin-$D_{98\%,CTV}$ Gy(RBE)
45.0 Gy(RBE)	27.9	46.8	47.9	43.9	42.8
50.4 Gy(RBE)	96.0 (39.7-340.1)	52.1 (51.7-52.9)	52.9 (52.4-53.4)	49.5 (49.1-50.0)	48.5 (47.8-49.0)
54.0 Gy(RBE)	182.3 (92.6-364.3)	55.7 (55.3-56.1)	56.6 (56.2-57.1)	52.8 (52.2-53.4)	51.4
59.4 Gy(RBE)	14.2 (14.0-14.4)	61.1 (60.4-61.7)	62.0 (61.3-62.8)	57.8 (54.3-58.9)	54.6 (51.8-57.1)
60.0 Gy(RBE)	47.2	63.5	65.6	58.0	53.6

Jesús Rojo-Santiago, Steven J.M. Habraken, D. Lathouwers et al.

Radiotherapy and Oncology 163 (2021) 121–127

Table 2
Clinical Goals on the CTV and OARs for the neuro patient cohort.

Structure	Criterion
CTV	VWmin-$D_{98\%,CTV} \geq 95\%$ D_{pres} Gy(RBE)
CTV	VWmax-$D_{2\%,CTV} \leq 107\%$ D_{pres} Gy(RBE)
Brainstem Core	VWmax-$D_{0.03cc}$ < 54 Gy(RBE)
Brainstem Surface	VWmax-$D_{0.03cc}$ < 60 Gy(RBE)
Spinal Cord Core	VWmax-$D_{0.03cc}$ < 54 Gy(RBE)
Spinal Cord Surface	VWmax-$D_{0.03cc}$ < 60 Gy(RBE)
Optical Chiasma	VWmax-$D_{0.03cc}$ < 55 Gy(RBE)
Cochlea (L+R)	VWmax-D_{mean} < 45 Gy(RBE)
Optical Nerves (L+R)	VWmax-$D_{0.03cc}$ < 55 Gy(RBE)
Retinae (L+R)	VWmax-$D_{0.03cc}$ < 45 Gy(RBE)
Lenses (L+R)	VWmax-$D_{0.03cc}$ < 10 Gy(RBE)
Pituitary	VWmax-D_{mean} < 20 Gy(RBE)
Lacrimal Glands (L+R)	VWmax-D_{mean} < 30 Gy(RBE)
Hippocampi	VWmax-D_{mean} < 7.4 Gy(RBE)

mean-square (RMS) value of the residual setup standard deviations, for all the patients [8]. Machine specifications were determined during acceptance and commissioning of the treatment chain, and are maintained through a regular QA program. This is further discussed in the Results section.

In our center, the CT-value to proton SPP calibration of the SECT acquisition and reconstruction protocol was based on measurements with a Gammex 467 phantom. It is known from literature [22–24] that in neuro-oncological treatments with SECT this leads to a systematic underestimation of proton SPP of 1.2% ± 0.7% (1 SD) [22], i.e., a range undershoot, as compared to the Dual-energy CT (DECT). We assumed an additional uncorrelated SPP error in DECT of 0.7% [23] and arrived at a total systematic SPP error of 1.2% ± 1.0% (1 SD).

Polynomial chaos expansion (PCE)

PCE, as implemented in Matlab (version R2017a [31]), provides a computationally fast, patient- and plan-specific model of the dependence of a 3D dose distribution on treatment uncertainties. The dose D_i to each voxel i, as affected by a geometrical shift $\vec{\xi} = (\xi_x, \xi_y, \xi_z)$ and a relative range error ρ, is approximated by the analytical series expansion $D_i(\vec{\xi}, \rho) = \sum_{k=0}^{P} a_{i,k} \Psi_k(\vec{\xi}, \rho)$, with expansion coefficients $\{a_{i,k}\}$ and multi-dimensional Hermite polynomials $\Psi_k(\vec{\xi}, \rho)$ [19,32]. The expansion coefficients are approximated by linear regression (i.e., linear in the coefficients ($a_{i,k}$), and the number of polynomial terms and regression points are both chosen such that an optimum between model accuracy and

computational time is achieved. The PCE models used here are based on 9 shifted scenarios (1 nominal scenario, 6 scenarios with setup errors of ± SR mm, and 2 scenarios with range errors of ± RR %) to select the voxels taken into account, and 208 scenarios to construct the model, using the clinical RayStation dose engine with a Monte Carlo (MC) uncertainty of 1%. For details we refer to the supplementary information [SM] and references [10,19].

After initial validation and PCE parameter optimization in several patients, more extensive and systematic validation of PCE for the present application was performed for a robust and a more challenging clinically non-robust plan from the cohort, including absolute dose differences, dose-volume histograms (DVH) comparisons and dosimetric parameters dependencies [SM].

PCE-based robustness evaluation

A PCE model of the dose was constructed for the clinical treatment plan for all 26 patients here. Using this model, 100.000 complete fractionated treatments were simulated for each patient, drawing (i) a fixed systematic setup and (ii) a fixed range error for the complete treatment and (iii) a random setup error for each treatment fraction, sampled from the Gaussian error distributions corresponding to the uncertainties listed in Table 3.

As the primary goal of robust optimization and evaluation is to ensure adequate target dose in the presence of errors, the resulting 100.000 dose distributions were evaluated to obtain probability histograms and patient-specific distributions of $D_{98\%,CTV}$ and $D_{2\%,CTV}$ [SM] across the sampled treatments courses (scenario distributions), normalized to the D_{pres} of each patient. Population-dose histograms are calculated by averaging (over all patients from the cohort) the patient-specific probabilities, as derived from the $D_{98\%,CTV}$ and $D_{2\%,CTV}$ scenario distributions, using the patient-population mean as a fair estimator of the actual population probability [SM].

Results

Results of residual patient setup data and machine specification are listed in Table 3. Similar systematic and random geometrical uncertainties were found for the lateral (patient left–right), the dorsoventral and the craniocaudal directions (1 SD error of 0.11, 0.15 and 0.09 mm for the systematic error and 0.24, 0.22 and 0.19 mm for random uncertainties respectively). Since no significant differences between the directions were found ($p > 0.2$), isotropic systematic and random error distributions, based on the RMS of the errors, were used. Geometrical systematic and random error distributions were centered around 0, while no significant

Table 3
Geometrical and range uncertainties taken into account for the PCE robustness evaluation. Uncertainties are split into a systematic and a random component.

Range uncertainties (%)						
Uncertainty	Systematic ± 1 SD					
SPP	1.2 ± 1.0					

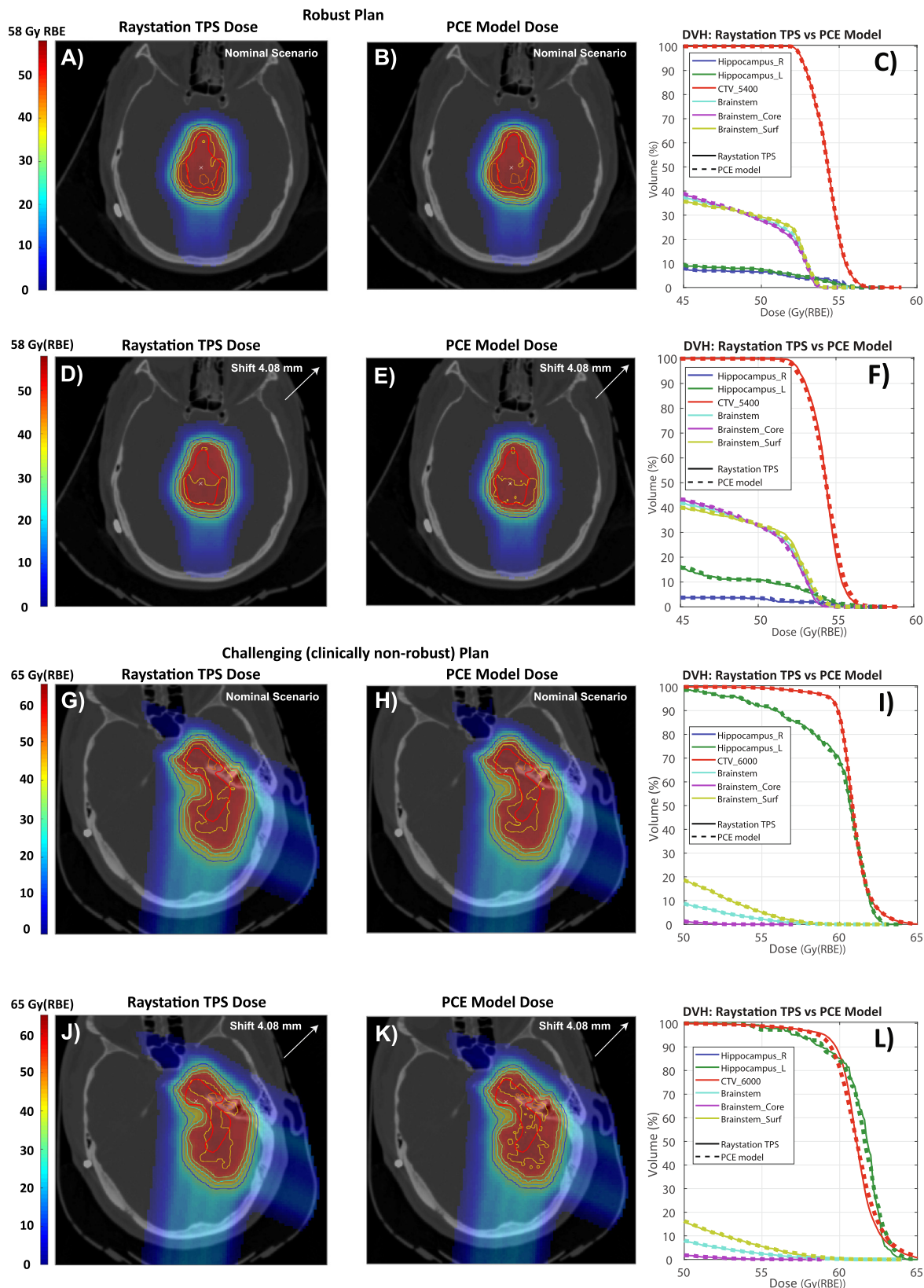| Geometrical uncertainties (mm)						
Uncertainty Isocentric inaccuracies	Left-right	Systematic ± 1 SD Dorsoventral	SD Craniocaudal	Left-right	Random ± 1 SD Dorsoventral	Craniocaudal
CT isocenter	±0.5	±0.5	±0.5			
Isocenter gantry				±0.5	±0.5	±0.5
Couch				±0.2	±0.2	±0.2
Online matching				±0.5	±0.5	±0.5
MR registration	±0.5	±0.5	±0.5			
Patient setup errors	Left-right	Dorsoventral	Craniocaudal	Left-right	Dorsoventral	Craniocaudal
Residual setup	0.09 ± 0.11	−0.01 ± 0.15	0.02 ± 0.09	0.09 ± 0.24	−0.01 ± 0.22	0.02 ± 0.19
Total geometrical	0.09 ± 0.72	−0.01 ± 0.72	0.02 ± 0.71	0.09 ± 0.77	−0.01 ± 0.77	0.02 ± 0.76
Geometrical error		0.0 ± 0.7			0.0 ± 0.8	

Fig. 1. PCE validation results for a robust (patient 4) and a more challenging (clinically non-robust, patient 8) treatment plan from the patient cohort, treated to 54 Gy (RBE) and 60 Gy (RBE) respectively, both in 30 fractions. From left to right, the panels show (i) an axial slice of the clinical RayStation TPS dose distribution, (ii) the PCE model of the same slice and (iii) CTV and OAR DHVs of the TPS and PCE dose distribution. From top to bottom, the nominal scenario (free of geometrical errors and a fixed overall SPP error of 1.2%) and a shift of 4.08 mm (3.58σ scenario, corresponding to a 91% percentile shift from the error distribution) for the clinically robust plan (from A) to F)) and a non-robust one (from G) to L)) respectively. All dose distributions were calculated for a fixed range undershoot of 1.2% [22–24]. The direction of the shifts is indicated by the white arrows.

Jesús Rojo-Santiago, Steven J.M. Habraken, D. Lathouwers et al.

Radiotherapy and Oncology 163 (2021) 121–127

deviations from 0 of the overall means (overall means < 1 SD) were found.

Errors in the registrations of (i) the MRI to planning-CT (pCT) and (ii) online orthogonal planar kV image pairs to digitally reconstructed radiographs (DRRs), derived from the pCT, also contribute to the total systematic and random geometrical error. Furthermore, isocenter alignment errors (i) at the pCT, (ii) of the treatment couch and (iii) in the online imaging relative to proton beam also led to systematic and random geometrical errors. We assumed that these errors are normally distributed and use the tolerance levels defined during acceptance as 1 SD errors. Total systematic and random geometrical contributions were calculated as the RMS of the errors per direction.

PCE validation results for a robust (from Fig. 1A to F) and a more challenging (clinically non-robust) plan (from Fig. 1G to L) are displayed in Fig. 1. The panels show the clinical RayStation TPS dose (left) and the PCE model dose (middle) in an axial slice through the center of the CTV. Both CTV and OAR DVHs (right) are displayed for the nominal scenario and for an isocenter shift of 4.08 mm, with a range uncertainty of 1.2% for the two treatment plan validations. For the nominal scenario, PCE agrees with the clinical TPS in less than 0.4% for both of the treatment plans. For the shifted scenario, PCE starts to show voxel dose differences of 1.1% and 2.3% for the robust and challenging plan respectively, still accurately approximating the relevant clinical DVH parameters ($D_{98\%,CTV}$, $D_{2\%,CTV}$) for the latter one. Additional PCE validation results are discussed in the supplementary material [SM].

Scenario distributions for $D_{98\%,CTV}$, $D_{50\%,CTV}$ and $D_{2\%,CTV}$ were derived for all the patients. Two typical examples are shown in Fig. 2. For the treatment plans in which the clinical criteria were met (clinically robust plans, cf., Fig. 2A), scenario $D_{98\%,CTV}$ distribution is typically left-skewed, showing small differences in $D_{98\%,CTV}$ values. A mean $D_{98\%,CTV}/D_{pres}$ value of 98.4% ± 0.6 % for the 98% of the sampled treatment courses (98% percentiles of the $D_{98\%,CTV}$ distributions) is found for this subgroup. On the other hand, treatment plans in which robust target dose was sacrificed (clinically non-robust plans, cf., Fig. 2B), showed a sizeable probability of not achieving the planning constraint, obtaining wider scenario $D_{98\%,CTV}$ distributions with treatment courses extending below the clinical criteria. No hot spots of dose were found for any treatment plan, as it can be seen in the $D_{2\%,CTV}$ scenario distribution in both subgroups.

Fig. 3A shows the PCE-based probability of achieving adequate target dose (i.e., $D_{98\%,CTV} \geq 95\% D_{pres}$) for all 26 patients, in decreasing order. Within the clinically robust subgroup of 20 patients,

indicated in blue, the average probability of achieving adequate CTV dose is 100%, within 0.01%. For the clinically non-robust plans, the probability of achieving adequate target dose was almost 0% in 3/6 SFUD plans, while values above 80% were realized in the other 3/6 cases (2 SFUD and 1 MFO plan). One case in which the clinical robustness evaluation criterion was not met (VWmin-$D_{98\%,CTV}$/D_{pres} = 0.94(9)[1]), did turn out to be robust in terms of the PCE-based metric (PCE-$D_{98\%CTV}/D_{pres}$ = 0.98).

The population-dose histogram ($D_{98\%,CTV}$) for the entire patient cohort was also derived, based on the population mean probability [SM] (Fig. 3B). The robustness constraint ($D_{98\%,CTV}/D_{pres}$ = 0.95) was met with a population probability of 97.7%, including clinically robust and non-robust treatment plans. In the entire patient population, the realized CTV dose, evaluated at a population probability of 98%, is on average 0.1 percentage points (p.p.) lower than prescribed (range −2.4 p.p. to +0.5 p.p.). For the clinically robust subgroup (20 out of 26 plans), the CTV dose is on average 3.0 p.p. higher than prescribed (range +2.7 p.p. to +3.5 p.p.). However, a lack of robustness ($D_{98\%,CTV}/D_{pres}$ < 0.95) is revealed for the clinically non-robust plans, with the exception of the borderline non-robust treatment plan for patient 21, with a CTV dose of 0.98, 0.94(7)[1], 0.94, 0.91, 0.90 and 0.89 for the patients 21 to 26 respectively. Furthermore, population $D_{2\%,CTV}$ dose histogram for this patient cohort is also included in the SM.

Discussion

To our best knowledge, the results presented here constitute the first clinical application of PCE, integrated with the MC dose engine of a TPS to assess robustness in clinical PT treatment plans. In comparison to other approaches [2,14–16], which are limited to a few hundred evaluation scenarios, the major advantage of PCE is that, using 208 scenario dose evaluations, it is feasible to simulate 100.000 complete fractionated treatments. This not only allows to obtain accurate statistics on clinically relevant dosimetric parameters, with a statistical error of about 0.3%, but also to obtain probabilistic robustness metrics (e.g., $D_{98\%,CTV}$, $D_{2\%,CTV}$) and scenario distributions. Depending on the target volume, the complete PCE robustness evaluation was achieved in a few of hours per patient, taking around 20% of the time to compute the 100.000 fully fractionated treatments with PCE (corresponding to computing from 2.8 to 3.3 million separate fraction dose distributions with up to 7.000.000 voxels) and 80% to analyze all treatment doses (adding up fraction doses, calculating DVHs, etc.). For a sample size of 1000 fractionated treatments, the robustness analysis would take
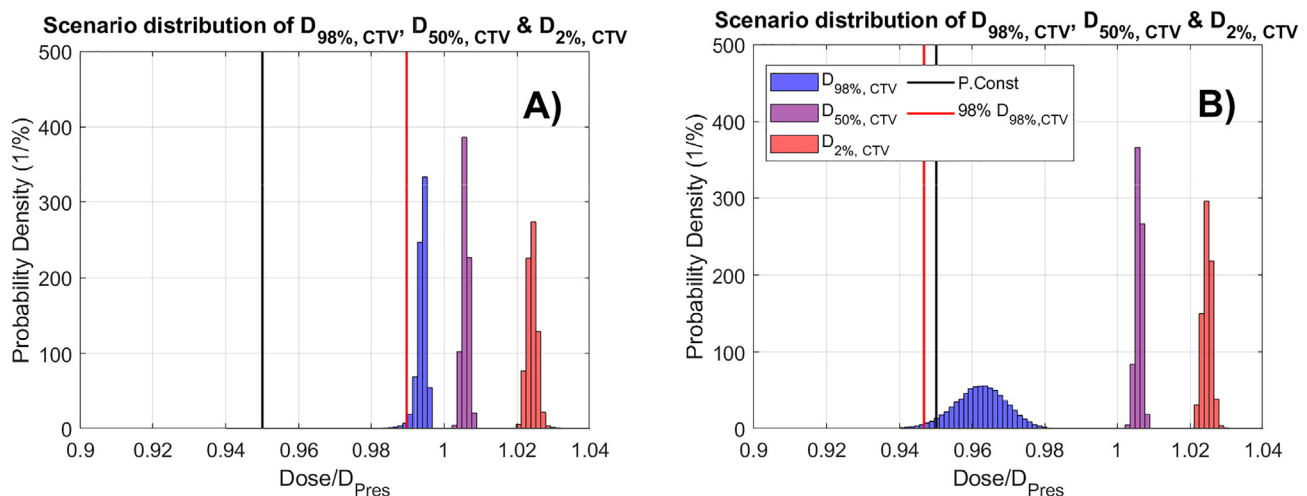

Fig. 2. $D_{98\%,CTV}$, $D_{50\%,CTV}$ and $D_{2\%,CTV}$ scenario distributions for a robust plan according clinical protocol (left, patient 3) and one in which robust target dose was sacrificed to spare critical OARs (right, patient 22). The planning constraint and 98% percentile values of the $D_{98\%,CTV}$ are also shown.
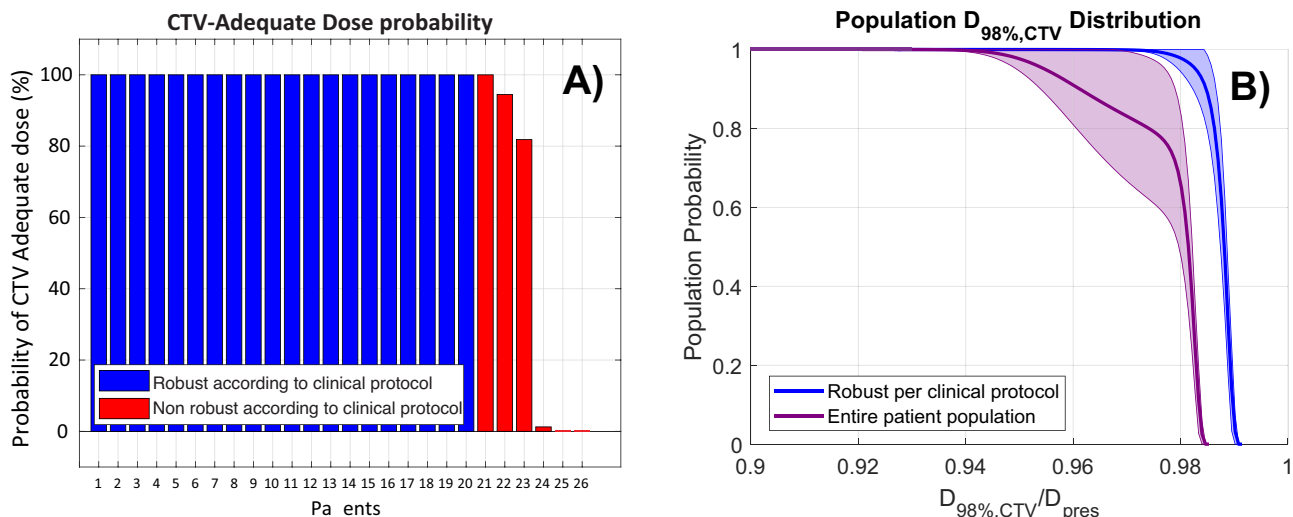
Fig. 3. PCE-based robustness evaluation results. (A) PCE probability of adequate target dose to clinically robust (red) and non-robust (blue) plans. (B) PCE population $D_{98\%,CTV}$ histogram with 98% confidence interval, regarding the entire patient cohort (purple) and robust (blue) plans according to clinical protocol. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

only a couple of minutes. This could be further reduced assuming infinitely fractionated treatments, sampling scenarios based on the RMS value of systematic and random errors.

Our main finding is that the clinical robustness evaluation protocol is safe in terms of CTV dose for this patient group since all plans that fulfill the clinical robustness criteria were found to be robust also in the PCE evaluation. Moreover, for treatment plans that were non-robust in the PCE-based evaluation, CTV dose was also lower than prescribed in the clinical robustness evaluation as CTV coverage was sacrificed to spare OARs. We found that the clinical protocol, using 3 mm setup and 3% range relative SPP robustness settings, is conservative for clinically robust treatment plans considering the error distributions measured at our center (Table 3). Based on 98% population coverage, CTV dose in these plans was on average 3.0% higher than prescribed in terms of the PCE-based $D_{98\%,CTV}$. The small variation (0.8 p.p in CTV dose at a 98% population probability) in the population-dose histogram for the clinically robust subgroup (Fig. 3B) indicates that the clinical robustness protocol, which uses 21 error scenarios for the optimization and 28 for the evaluation, is sufficient to obtain consistent results. We did not find any significant differences in robustness between the 19 MFO and 7 SFUD treatment plans. The clinical robustness protocol is conservative in terms of target dose according to the robustness recipes in reference [30] and the clinical patient setup data presented in Table 3, which would justify a setup robustness setting of 2.7 mm. From this we conclude there may be room for further optimization of the robustness settings and at our center, a reduction of the setup robust optimization setting might be appropriate for this patient group. To this end, a PCE-based robustness evaluation with a set of treatment plans with different robustness settings is required.

Our population-based error analysis and PCE methodology give important insights in the performance of the Dutch consensus protocol for this treatment site, including a probabilistic interpretation of the VWmin-$D_{98\%,CTV}$ and VWmax-$D_{2\%,CTV}$ values. As such, our results are relevant to the entire patient population more than dose accumulation for individual patients would be.

As PCE is an analytical approximation of the dose engine, its validation is essential. By construction, it deviates from the clinical TPS dose for sufficiently large errors. The PCE model parameters and the TPS MC uncertainty used here were chosen such that, for isocentric shifts up to 4 mm, the PCE dose distributions agree

within 1% with the clinical TPS doses. For other treatment sites, other dosimetric endpoints or in case of bigger uncertainties, a higher PCE accuracy may be desirable. This could be achieved by increasing the polynomial order of the expansion and/or the accuracy of the numerical integration or regression methods to determine the expansion coefficients [19], increasing the number of TPS dose scenarios required to build the PCE and, hence, an increased calculation time. In calculating the expansion coefficients, we used a regression approach [33], which minimizes the impact of TPS MC uncertainty [SM]. To achieve a better model accuracy, also a lower TPS MC uncertainty could be necessary, leading to an increased computational time.

Our analysis is limited to isocentric errors, modelled as rigid shifts, and relative SPP errors. In our center, patient rotations from the treatment beam and the CT couch are corrected by a 6D robotic treatment couch, but nonzero systematic and random residual patient rotation errors of the order of 0.2° remain. They have the biggest impact at the beam entrance, and for a typical distance of 4 cm between the center of CTV and the beam entrance, correspond to a displacement on the order of 0.1 mm. This is considered negligible in comparison to the errors taken into account here. The analysis of post-fraction kV image pairs shows that intrafraction motion [34] is small in most patients and does not correlate with residual pre-treatment setup errors [SM]. It could be taken into account as another, uncorrelated, random patient setup error but would have little impact on the results. As post-fraction data is polluted by some outliers, likely due to patients being aware of the end of proton beam delivery, it was not taken into account here.

Another source of geometrical uncertainty not taken into account here, are anatomical variations over the course of treatment. In our center, these are mitigated by a plan adaptation protocol. Based on the evaluation of weekly cone-beam CT scans, a new pCT and treatment plan is made when necessary. Significant anatomical changes and, hence, plan adaptations are rare in patients with neurological tumors.

Conclusion

We have performed an advanced analysis of the setup and range (SPP) robustness in a cohort of 27 clinical neuro-oncological PT treatment plans, 26 of which could be evaluated by simulating realistic errors in 100.000 fractionated treatments

Jesús Rojo-Santiago, Steven J.M. Habraken, D. Lathouwers et al.

Radiotherapy and Oncology 163 (2021) 121–127

for each patient. The clinical protocol, using 3 mm setup and 3% SPP robustness settings in treatment planning, is safe in terms of CTV dose for this patient group considering the error distributions measured at our center. For the treatment plans that meet the clinical robustness constraints, it is conservative in terms of CTV dose and leads to little variation between patients. In view of the computational advantage of PCE, as compared to repeated TPS MC scenario dose calculations, it is feasible to perform an automated PCE validation and a PCE-based robustness evaluation overnight. It could complement, or, in the longer run, replace, the current clinical robustness evaluation protocol.

Ethics approval

The data used in this study originates from the research database of HollandPTC. This database consists of data from all consenting patients treated in HollandPTC. The local Institutional Review Board of LUMC waived the need to assess the protocol of the research database.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2021.07.028.

References

[1] Unkelbach J, Bortfeld T, Martin BC, Soukup M. Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning. Med Phys 2009;36:149–63.

[2] Albertini F, Hug EB, Lomax AJ. Is it necessary to plan with safety margins for actively scanned proton therapy? Phys Med Biol 2011;56:4399.

[3] Lomax AJ. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: The potential effects of inter-fraction and inter-field motions. Phys Med Biol 2008;53:1043–56.

[4] Unkelbach J, Chan TCY, Bortfeld T. Accounting for range uncertainties in the optimization of intensity modulated proton therapy. Phys Med Biol 2007;52:2755–73.

[5] Fredriksson A, Bokrantz R. A critical evaluation of worst-case optimization methods for robust intensity-modulated proton therapy planning. Med Phys 2014;41:081701.

[6] Fredriksson A, Forsgren A, Hårdemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. Med Phys 2011;38:1672–84.

[7] Chen W, Unkelbach J, Trofimov A, Madden T, Kooy H, Bortfeld T, et al. Including robustness in multi-criteria optimization for intensity-modulated proton therapy. Phys Med Biol 2012;57:591–608.

[8] Van Herk M. Errors and margins in radiotherapy. 2004;14:52–64.

[9] van Herk M, Remeijer P, Rasch C, Lebesque JV. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. Int J Radiat Oncol Biol Phys 2000;47:1121–35.

[10] van der Voort S, van de Water S, Perkó Z, Heijmen B, Lathouwers D, Hoogeman M. Robustness recipes for minimax robust optimization in intensity

[11] modulated proton therapy for oropharyngeal cancer patients. Int J Radiat Oncol Biol Phys 2016;95:163–70.

[11] Lomax AJ. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties. Phys Med Biol 2008;53:1027.

[12] McGowan SE, Burnet NG, Lomax AJ. Treatment planning optimisation in proton therapy. Br J Radiol, 2013;86:20120288.

[13] Hernandez V, Hansen CR, Widesott L, Bäck A, Canters R, Fusella M, et al. What is plan quality in radiotherapy? The importance of evaluating dose metrics, complexity, and robustness of treatment plans. Radiother Oncol 2020;153:26–33.

[14] Henríquez FC, Castrillón SV. A novel method for the evaluation of uncertainty in dose–volume histogram computation. Int J Radiat Oncol Biol Phys 2008;70:1263–71.

[15] Park PC, Cheung JP, Zhu XR, Lee AK, Sahoo N, Tucker SL, et al. Statistical assessment of proton treatment plans under setup and range uncertainties. Int J Radiat Oncol Biol Phys 2013;86:1007–13.

[16] Buti G, Souris K, Barragán Montero AM, Cohilis M, Lee JA, Sterpin E. Accelerated robust optimization algorithm for proton therapy treatment planning. Med Phys 2020;47:2746–54.

[17] Korevaar EW, Habraken SJM, Scandurra D, Kierkels RGJ, Unipan M, Eenink MGC, et al. Practical robustness evaluation in radiotherapy–A photon and proton-proof alternative to PTV-based plan evaluation. Radiother Oncol 2019;141:267–74.

[18] Lomax AJ, Pedroni E, Rutz HP, Goitein G. The clinical potential of intensity modulated proton therapy. Zeitschrift für Medizinische Physik 2004;14:147–52.

[19] Perkó Z, van der Voort SR, van de Water S, Hartman CMH, Hoogeman M, Lathouwers D. Fast and accurate sensitivity analysis of IMPT treatment plans using Polynomial Chaos Expansion. Phys Med Biol 2016;61:4646–64.

[20] Nguyen TTC, Nguyen BT, Mai NV. Robustness evaluation of intensity modulated proton therapy plans using dose volume population histogram. Physica Med 2019;65:219–26.

[21] Trofimov A, Unkelbach J, DeLaney TF, Bortfeld T. Visualization of a variety of possible dosimetric outcomes in radiation therapy using dose-volume histogram bands. Pract Radiat Oncol 2012;2:164–71.

[22] Wohlfahrt P, Möhler C, Troost EGC, Greilich S, Richter C. Dual-energy computed tomography to assess intra-and inter-patient tissue variability for proton treatment planning of patients with brain tumor. Int J Radiat Oncol Biol Phys 2019;105:504–13.

[23] Wohlfahrt P, Möhler C, Richter C, Greilich S. Evaluation of stopping-power prediction by dual-and single-energy computed tomography in an anthropomorphic ground-truth phantom. Int J Radiat Oncol Biol Phys 2018;100:244–53.

[24] Wohlfahrt P, Möhler C, Stützer K, Greilich S, Richter C. Dual-energy CT based proton range prediction in head and pelvic tumor patients. Radiother Oncol 2017;125:526–33.

[25] Weide HLVD, Kramer MCA, Scandurra D, et al. Proton therapy for selected low grade glioma patients in the Netherlands. Radiother Oncol 2020.

[26] ICRU. Prescribing, recording, and reporting photon beam therapy (Report 50). 1993.

[27] ICRU. Report 62: Prescribing, recording and reporting photon beam therapy (Suppl to 50). J ICRU 1999.

[28] Eekers DBP, in 't Ven L, Roelofs E, Postma A, Alapetite C, Burnet NG, et al. The EPTN consensus-based atlas for CT-and MR-based contouring in neuro-oncology. Radiother Oncol 2018;128:37–43.

[29] Eekers DBP, Di Perri D, Roelofs E, et al. Update of the EPTN atlas for CT- and MR-based contouring in Neuro-Oncology. 2021;160:259–265.

[30] Ter Haar C, Habraken S, Lathouwers D, Wiggenraad R, Krol S, Perkó Z, et al. PO-0998 setup and range robustness recipes for skullbase meningioma IMPT using Polynomial Chaos Expansion. Radiother Oncol 2019;133:S549–50.

[31] Perkó Z. Open source generalized Polynomial Chaos Expansion (openGPC) Toolbox. https://gitlab.com/zperko/opengpc.

[32] Le Maître OP, Knio OM. Spectral methods for uncertainty quantification: with applications to computational fluid dynamics. Dordrecht, The Netherlands. 2010.

[33] Perkó Z. Sensitivity and Uncertainty Analysis of Coupled Reactor Physics Problems: Method Development for Multi-Physics in Reactors. Technological University of Delft; 2015.

[34] Grassberger C, Dowdell S, Lomax A, et al. Motion interplay as a function of patient parameters and spot size in spot scanning proton therapy for lung cancer. Int J Radiat Oncol Biol Phys 2013;86:380–6.