Physics and Imaging in Radiation Oncology 20 (2021) 117–120

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/physics-and-imaging-in-radiation-oncology

Technical Note

Combining inter-observer variability, range and setup uncertainty in a variance-based sensitivity analysis for proton therapy

Jan Hofmaier [a,*], Franziska Walter [a], Indrawati Hadi [a], Maya Rottler [a], Rieke von Bestenbostel [a], George Dedes [c], Katia Parodi [c], Maximilian Niyazi [a], Claus Belka [a,d], Florian Kamp [a,b]

[a] Department of Radiation Oncology, University Hospital, LMU Munich, Munich, Germany
[b] Department of Radiation Oncology and CyberKnife Center, Faculty of Medicine, University Hospital Cologne, Cologne, Germany
[c] Department of Medical Physics, Faculty of Physics, LMU Munich, Munich, Germany
[d] German Cancer Consortium (DKTK), Munich, Germany

ARTICLE INFO

Keywords:
Proton therapy
Range uncertainty
Setup uncertainty
Inter-observer variability
Sensitivity analysis

ABSTRACT

Margin concepts in proton therapy aim to ensure full dose coverage of the clinical target volume (CTV) in presence of setup and range uncertainty. Due to inter-observer variability (IOV), the CTV itself is uncertain. We present a framework to evaluate the combined impact of IOV, setup and range uncertainty in a variance-based sensitivity analysis (SA). For ten patients with skull base meningioma, the mean calculation time to perform the SA including 1.6×10^4 dose recalculations was 59 min. For two patients in this dataset, IOV had a relevant impact on the estimated CTV $D_{95\%}$ uncertainty.

1. Introduction

Treatment plans in proton therapy are affected by range and setup uncertainties. These are typically compensated through margin concepts or robust planning approaches. Margin concepts aim at covering the clinical target volume (CTV) in presence of range and setup uncertainty [1]. However, due to inter-observer variability (IOV), the CTV itself is uncertain. While there are many studies assessing IOV, only few studies have investigated dosimetric consequences of IOV [2], e.g Lobefalo et al. [3] who investigated the dosimetric impact of IOV in three-dimensional conformal radiotherapy (3D-CRT) and volumetric modulated arc therapy for rectal tumours, Hellebust et. al. [4] who assessed the dosimetric impact of IOV in brachytherapy for cervical cancer and Eminowicz et al. [5], who studied the dosimetric impact of IOV in VMAT for cervical cancer. To the best of our knowledge, there is no study assessing the combined and relative impact of range, setup uncertainty and IOV in proton therapy in a quantitative way. The statistical method of variance-based sensitivity analysis (SA) is suited for this, since it can be used to assess the impact of uncertainty of multiple input parameters on the output of a quantitative model [6]. In the context of patient dose calculation in medical physics, the technique has been previously applied to relative biological effectiveness (RBE) uncertainties in carbon ion therapy [7,8] and to estimate the impact of interpatient variability on organ dose estimates in nuclear medicine [9]. Recently, a framework to evaluate the combined impact of range, setup and RBE uncertainty in a variance-based SA has been presented by our group [10]. In this technical note, an extension of the framework to include IOV is shown. The feasibility of the approach was demonstrated by using it to investigate the relative impact of IOV, range and setup uncertainty on proton plans for a dataset with ten patients with skull base meningioma.

2. Materials and methods

2.1. Variance-based sensitivity analysis

In the Monte Carlo method of global variance-based SA, the output of a model $Y = f(X)$ with k input factors $X = (x_1, x_2, \ldots, x_k)$ which are subject to uncertainty is recalculated many times while simultaneously and randomly varying the input factors within their assumed distributions. In our particular case, the model $f(X)$ corresponded to a dose calculation followed by a dose volume histogram (DVH) calculation. The output Y corresponded to DVH parameters of interest. The input factors $(x_1, x_2 \ldots x_k)$ included patient shifts in three spatial dimensions, absolute and relative range shifts as well as IOV, resulting in $k = 6$ input factors. The resulting variance $V(Y)$ is decomposed as [6]:

* Corresponding author.
E-mail address: jan.hofmaier@med.uni-muenchen.de (J. Hofmaier).

https://doi.org/10.1016/j.phro.2021.11.005
Received 30 September 2021; Received in revised form 16 November 2021; Accepted 16 November 2021

$$V(Y) = \sum_{l=1}^{k} V_l + \sum_{l=1}^{k}\sum_{m>l}^{k} V_{lm} + \sum_{l=1}^{k}\sum_{m>l}^{k}\sum_{n>m}^{k} V_{lmn} + \ldots + V_{1\ldots k} \quad (1)$$

resulting in $(2^k - 1)$ terms. The first order terms are

$$V_l = V[E(Y|X_l)] \quad (2)$$

The expectation value $E(Y|X_l)$ is hereby calculated over all possible values of all input factors except for X_l, which is kept fixed. The second order terms, which are representing the interaction between the inputs X_l and X_m, are

$$V_{lm} = V[E(Y|X_l, X_m)] - V_l - V_m \quad (3)$$

Higher order terms are defined in an analoguous fashion. Sensitivity indices are defined by normalising to the overall variance

$$S_l = \frac{V_l}{V(Y)} \quad (4)$$

$$S_{lm} = \frac{V_{lm}}{V(Y)} \quad (5)$$

and so on. Total effect indices are defined by summing all terms of any order containing l:

$$ST_l = S_l + \sum_{m\neq l}^{k} S_{lm} + \ldots + S_{1\ldots k} \quad (6)$$

Like in a previous study from our group [10], the efficient Monte Carlo method proposed by Saltelli [6] was used for direct calculation of S_l and ST_l, and sampling from low-discrepancy quasi-random sequences was employed to improve convergence. This method requires $N(k+2)$ model evaluations, where N is typically of the order of 10^3. In our study, as described above, we had $k = 6$ input factors. We set $N = 2048$, which resulted in approximately $1.6{\cdot}10^4$ model evaluations. The sensitivity analysis framework was extended to include IOV. Additionally to the fast, graphics processing unit (GPU) based pencil beam algorithm capable of modeling setup and range variations described in the previous publication from our group [10], the possibility to include multiple treatment plans and to switch randomly between them was added.

2.2. Clinical dataset

Datasets of ten patients with benign (WHO grade I) meningioma of the skull base were included in this study. For all patients, contrast enhanced magnetic resonance imaging (MRI) and DOTATATE positron emission tomography (PET) images were available in addition to a planning computed tomography (CT).

2.3. Target delineation and treatment planning

A rigid image registration of MRI, PET and planning CT images was performed. For each patient, four clinicians independently delineated the gross tumor volume (GTV) taking into account all imaging modalities ($GTV_{observer}$). A consensus GTV (GTV_{STAPLE}) was created using the simultaneous truth and performance level estimation (STAPLE) algorithm [11] in the research treatment planning system computational environment for radiological research (CERR) [12]. This implementation of an expectation-maximization algorithm generates a probabilistic estimate of the true volume based on the volumes delineated by multiple observers. The GTV_{STAPLE} was used as the "ground truth" GTV. As an example, the four $GTV_{observer}$ and the GTV_{STAPLE} contours for patient number 1 are shown in the supplementary material. The $CTV_{observer}$ and the CTV_{STAPLE} were defined as the respective GTV without any margins applied (i.e. GTV = CTV), as suggested in a current guideline [13]. To obtain the planning target volumes (PTVs), gantry-angle specific margins were applied. To compensate for proton range uncertainty, larger margins were applied in beam direction than laterally. The applied margins were 6, 5 and 3 mm in distal, proximal and lateral directions, respectively. For a typical margin receipe of 3.5% + 3 mm, the distal margin of 6 mm would correspond to a radiological target depth of approximately 9 cm. Since all tumours were at the skull base and therefore at similar depths, the same absolute margins were applied to all patients for simplicity. For each $CTV_{observer}$ a $PTV_{observer}$ was created. For each $PTV_{observer}$ of each patient a spot scanning proton treatment plan with one beam was generated using non-robust optimization, resulting in a total number of 40 treatment plans (four treatment plans for each of the ten patients). The gantry angle was chosen individually for each patient. The proton plans were optimized to deliver 1.8 Gy (RBE) per fraction to the $PTV_{observer}$. A spatially constant RBE of 1.1 was assumed.

2.4. Application of the SA framework

Like in the previous study from our group [10], the variance-based SA was performed assuming the following uncertainty distributions for the input factors mentioned in Section 2.1: For patient shifts in X,Y and Z directions, a normal distribution with standard deviation $\sigma_{X,Y,Z} =$ 1 mm truncated to $2\sigma_{X,Y,Z}$ was assumed. For relative range shifts the probability density was set to a normal distribution with standard deviation $\sigma_{r,rel} = 3$ % truncated to $2\sigma_{r,rel}$. Additionally, absolute range shifts following a normal distribution with standard deviation $\sigma_{r,abs} = 1$ mm truncated to $2\sigma_{r,abs}$ were assumed. For IOV, an equal probability of p = 0.25 for each of the four observer treatment plans was chosen. To perform the SA, the dose distribution was re-calculated approximately $1.6{\cdot}10^4$ times (corresponding to $N = 2048$ and $k = 6$ in the Saltelli formalism, as described in Section 2.1) while simultaneously sampling from the above uncertainty distributions. An Nvidia Quadro RTX 8000 GPU with 48 gigabytes of memory was used. For the resulting dose distributions, DVHs were calculated for the CTV_{STAPLE}. Confidence intervals (CIs) and sensitivity indices for the dose level enclosing 95% of the CTV_{STAPLE} ($D_{95\%}$) were calculated. Convergence plots of the sensitivity indices were created. The obtained total effect indices ST were converted to SI_{IOV}, the sum of all interaction terms with involvement of IOV and SI_{other}, the sum of all interaction terms without involvement of IOV. By definition is

$$SI_{IOV} = ST_{IOV} - S_{IOV} \quad (7)$$

and due to normalization

$$SI_{other} = 1 - S_{setup} - S_{range} - ST_{IOV} \quad (8)$$

3. Results

The mean calculation time to perform the $1.6{\cdot}10^4$ dose calculations was 59 min. Large differences were observed for the calculation times for different patients, which ranged from 11 min to 195 min. Convergence plots for S_l and ST_l for an exemplary patient are shown in panels A and B of Fig. 1. By visual inspection of the convergence plots it becomes evident that a sufficient convergence was achieved well below $N =$ 2048.

Results for the $D_{95\%}$ are presented in Table 1. For six patients, the width of the $CI_{95\%}$ for the $D_{95\%}$ was below 0.18 Gy (10% of the prescribed dose of 1.8 Gy). Uncertainties of more than 10 % were observed for patients 2, 3, 7 and 9. Here the widths of the $CI_{95\%}$ for the $D_{95\%}$ were 0.57, 0.24, 0.28 and 0.48 Gy, respectively. Plots of the DVHs for the CTV_{STAPLE} for these four patients with their corresponding 95 % and 68 % CIs are shown in panels C to F of Fig. 1. For two of these patients, the overall influence of IOV was negligible ($S_{IOV} + SI_{IOV} < 0.05$ for patients 7 and 9). In both cases, range uncertainty was the most important contribution to overall uncertainty (S_{range} was 0.53 and 0.70 for patients 7 and 9, respectively). For patients 2 and 3, however, IOV played a major

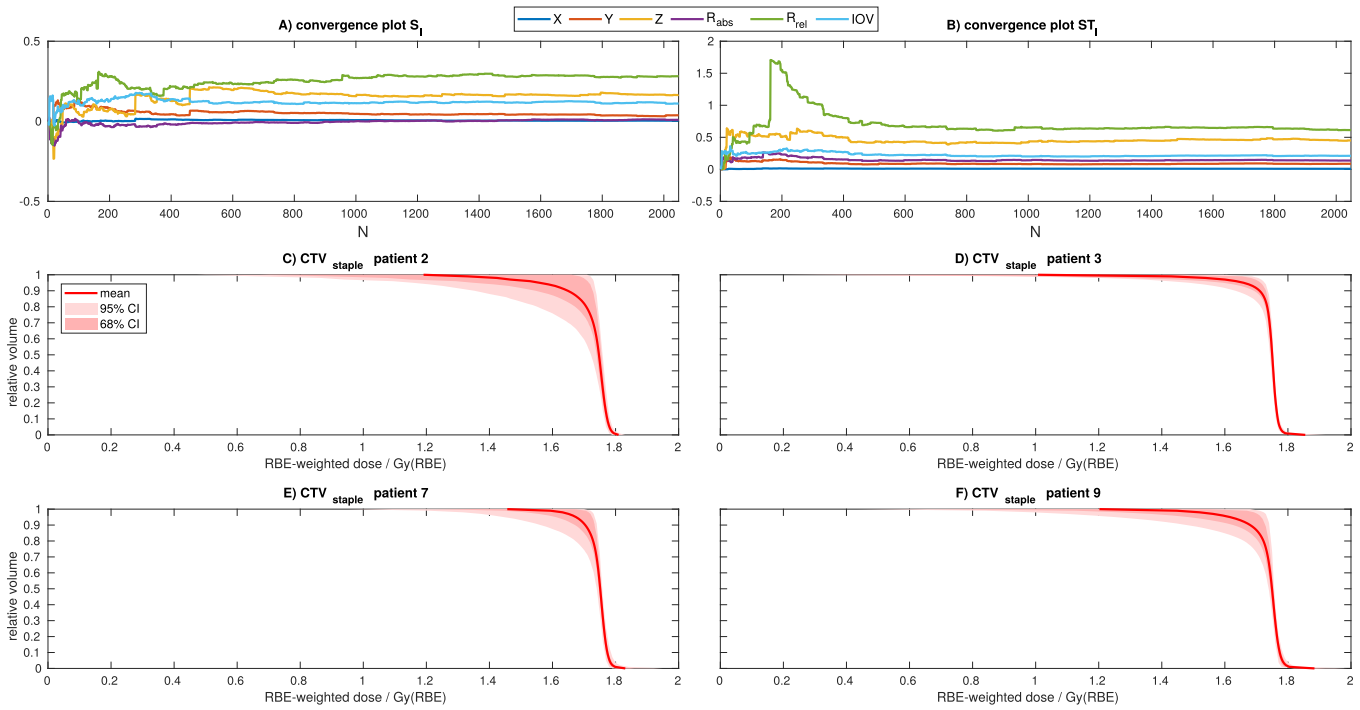

Fig. 1. Convergence plots for S_I and ST_I for one patient (panels A and B) and DVHs for CTV_{STAPLE} for the four patients with the largest overall $D_{95\%}$ uncertainties (panels C to F). The variability of the DVH in presence of setup uncertainty, range uncertainty and IOV is visualized by the shaded areas (68% and 95% CIs). The solid line indicates the mean value over all simulated error scenarios.

Table 1
Uncertainty and sensitivity analysis results for the $D_{95\%}$ for $(CTV)_{STAPLE}$. For each patient, the mean value and 95% and the 68% CIs have been calculated. The relative contribution to the overall uncertainty is broken down to first order indices S_{setup}, S_{range} and S_{IOV}, higher order indices with involvement of IOV (SI_{IOV}) and higher order indices without involvement of IOV (SI_{other}).

pat.	mean [Gy]	$CI_{95\%}$ [Gy]	$CI_{68\%}$ [Gy]	S_{setup}	S_{range}	S_{IOV}	SI_{IOV}	SI_{other}
1	1.71	1.62–1.74	1.69–1.73	0.21	0.29	0.11	0.10	0.29
2	1.56	1.16–1.73	1.35–1.72	0.12	0.35	0.34	0.09	0.10
3	1.66	1.49–1.73	1.59–1.71	0.14	0.11	0.62	0.01	0.12
4	1.70	1.59–1.73	1.68–1.72	0.19	0.14	0.10	0.12	0.45
5	1.73	1.71–1.74	1.72–1.74	0.37	0.31	0.00	0.14	0.18
6	1.73	1.70–1.74	1.72–1.73	0.20	0.29	0.03	0.04	0.44
7	1.67	1.45–1.73	1.61–1.73	0.25	0.53	0.00	0.02	0.20
8	1.69	1.58–1.74	1.65–1.73	0.23	0.49	0.04	0.05	0.19
9	1.62	1.25–1.73	1.48–1.71	0.13	0.70	0.01	0.00	0.16
10	1.73	1.70–1.74	1.72–1.74	0.26	0.24	0.02	0.06	0.42

role for overall uncertainty ($S_{IOV} + SI_{IOV}$ was 0.43 and 0.63 for patients 2 and 3, respectively).

4. Discussion

A framework for the variance-based SA of setup, range and IOV has been presented. To the best of our knowledge, this study is the first to assess the relative dosimetric impact of setup uncertainty, range uncertainty and IOV in a variance-based SA. In a first analysis of ten patients, calculation times were of the order of a few minutes to a few hours. These calculation times are fast enough for offline plan evaluation. Although this was not investigated in this study, it can be assumed that the differences in calculation time were caused by differences in the sizes and depths of the target volumes. The convergence plots in Fig. 1 suggest that actually less than $N = 2048$ would have been sufficient to achieve convergence, therefore the calculation times could be reduced by stopping the calculation after reaching a predefined convergence criterion. While for the majority of patients, the overall uncertainties in CTV coverage were small, in some cases the coverage was deteriorated. The dominating contributions to overall uncertainty were either range uncertainty or IOV. This suggests that IOV might have a relevant effect on target coverage in some patients.

In this work, the analysis was restricted to skull base meningioma, since the framework does not support organ motion at the moment. Furthermore, a pencil beam algorithm was used, whose accuracy is known to decrease in regions of high heterogeneity. The framework would be applicable without modification to other tumour sites for which these limitations are acceptable. The possibility to model motion could be included by extending the framework to use multiple CT geometries (e.g. phases of a 4D-CT to model breathing motion), at the cost of an increased memory usage and longer calculation times. In the previous publication from our group [10], uncertainties in variable RBE models were evaluated in combination with setup and range uncertainty. In this study, since the focus was on IOV, RBE uncertainty was not taken into account and a constant RBE of 1.1 was assumed. However, the combined evaluation of all four types of uncertainty could in principle also be included in the framework. This could be used in future studies to assess the combined impact of range, setup and RBE uncertainty and IOV. The evaluation of the CTV $D_{95\%}$ in presence of IOV required a "ground truth" CTV. Unfortunately, this volume is not known. In this

work, the consensus target volume created with the STAPLE algorithm was used to define a "ground truth" target volume, as has been done previously [14]. Since this algorithm provides a maximum likelyhood estimate for the actual CTV based on the observer CTVs themselves, this approach is well suited to capture the variability within a group of observers. However, it cannot correct systematic deviations from the ground truth CTV within the observer group. Furthermore, in this study only data from four observers was available, which was considered sufficient to show the feasibility of the approach. However, outlier contours could have considerable effect on the evaluation. For this reason, both the number of patients and the number of observers needs to be increased for future systematic evaluations of the impact of IOV in combination with setup and range uncertainties. Another limitation is that in our study simple proton plans with only one beam direction were used. More clinically realistic plans with multiple beam directions are supported by the framework without modifications, but have higher memory requirements and will lead to longer calculation times.

In this technical note, no metrics of contour similarity such as Dice coefficients or Hausdorff distances were evaluated. The presented framework might be used in future studies to investigate the correlation of these metrics with dosimetric parameters. It could also have potential applications in the investigation of the implications of uncertainty reduction. If technical advances such as dual energy computed tomography (DECT), proton CT and improved image guidance reduce range and setup uncertainty, the relative impact of IOV on overall uncertainty becomes larger. The SA framework could complement studies such as [15–17], who have investigated the impact of range and setup margin reduction. By also including IOV into the analysis, questions such as how far the overall uncertainty can be reduced by reduction of setup and range uncertainty before IOV becomes the limiting factor could be comprehensively investigated in future studies. Similarly, the following question could be assessed: Although not explicitly accounted for in the PTV concept, it can be assumed that IOV is compensated by the margins (or, in an analogous manner in the case of robust optimization, the plan robustness settings) to a certain extent. The SA framework could help to investigate whether a CTV-to-PTV margin reduction (or reduction of plan robustness settings) justified by reduced range and setup uncertainties would lead to an unexpected increase in uncertainty of CTV coverage caused by IOV.

In conclusion, a previously presented framework for variance-based sensitivity analysis has been extended to include IOV. The approach is feasible and enables the evaluation of the combined impact of setup and range uncertainty and IOV. In a first analysis of ten patients, IOV had a relevant impact on the CTV $D_{95\%}$ for two of these patients. This suggests that IOV could have a deteriorating effect on CTV coverage in some cases.

Declaration of Competing Interest

CB holds research grants from Elekta, Brainlab, Viewray and C-Rad without any relation to the research described here. The remaining authors declare that they have no competing interest.

Acknowledgements

This project was supported by the DFG grant KA 4346/1-1 and the DFG Cluster of Excellence Munich Center for Advanced Photonics (MAP).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.phro.2021.11.005.

References

[1] Unkelbach J, Alber M, Bangert M, Bokrantz R, Chan TCY, Deasy JO, et al. Robust radiotherapy planning. Phys Med Biol 2018;63(22):22TR02. https://doi.org/10.1088/1361-6560/aae659.
[2] Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. Radiother Oncol 2016;121(2):169–79. https://doi.org/10.1016/j.radonc.2016.09.009.
[3] Lobefalo F, Bignardi M, Reggiori G, Tozzi A, Tomatis S, Alongi F, et al. Dosimetric impact of inter-observer variability for 3D conformal radiotherapy and volumetric modulated arc therapy: The rectal tumor target definition case. Radiat Oncol 2013;8(1):176. https://doi.org/10.1186/1748-717X-8-176.
[4] Hellebust TP, Tanderup K, Lervåg C, Fidarova E, Berger D, Malinen E, et al. Dosimetric impact of interobserver variability in MRI-based delineation for cervical cancer brachytherapy. Radiother Oncol 2013;107(1):13–9. https://doi.org/10.1016/j.radonc.2012.12.017.
[5] Eminowicz G, Rompokos V, Stacey C, McCormack M. The dosimetric impact of target volume delineation variation for cervical cancer radiotherapy. Radiother Oncol 2016;120(3):493–9. https://doi.org/10.1016/j.radonc.2016.04.028.
[6] Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput Phys Commun 2010;181(2):259–70. https://doi.org/10.1016/j.cpc.2009.09.018.
[7] Kamp F, Brüningk S, Cabal G, Mairani A, Parodi K, Wilkens JJ. Variance-based sensitivity analysis of biological uncertainties in carbon ion therapy. Phys Med 2014;30(5):583–7. https://doi.org/10.1016/j.ejmp.2014.04.008.
[8] Kamp F, Wilkens JJ. Application of variance-based uncertainty and sensitivity analysis to biological modeling in carbon ion treatment plans. Med Phys 2019;46(2):437–47. https://doi.org/10.1002/mp.13306.
[9] Zvereva A, Kamp F, Schlattl H, Zankl M, Parodi K. Impact of interpatient variability on organ dose estimates according to MIRD schema: Uncertainty and variance-based sensitivity analysis. Med Phys 2018;45(7):3391–403. https://doi.org/10.1002/mp.12984.
[10] Hofmaier J, Dedes G, Carlson DJ, Parodi K, Belka C, Kamp F. Variance-based sensitivity analysis for uncertainties in proton therapy: A framework to assess the effect of simultaneous uncertainties in range, positioning and RBE model predictions on RBE-weighted dose distributions. Med Phys 2021;48(2):805–18. https://doi.org/10.1002/mp.14596.
[11] Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. IEEE Trans Med Imaging 2004;23(7):903–21. https://doi.org/10.1109/TMI.2004.828354.
[12] Deasy JO, Blanco AI, Clark VH. CERR: A computational environment for radiotherapy research. Med Phys 2003;30(5):979–85. https://doi.org/10.1118/1.1568978.
[13] Combs SE, Baumert BG, Bendszus M, Bozzao A, Brada M, Fariselli L, et al. ESTRO ACROP guideline for target volume delineation of skull base tumors. Radiother Oncol 2021;156:80–94. https://doi.org/10.1016/j.radonc.2020.11.014.
[14] Kristensen I, Nilsson K, Agrup M, Belfrage K, Embring A, Haugen H, et al. A dose based approach for evaluation of inter-observer variations in target delineation. Tech Innov Patient Support Radiat Oncol 2017;3-4:41–7. https://doi.org/10.1016/j.tipsro.2017.10.002.
[15] van de Water S, van Dam I, Schaart DR, Al-Mamgani A, Heijmen BJM, Hoogeman MS. The price of robustness; impact of worst-case optimization on organ-at-risk dose and complication probability in intensity-modulated proton therapy for oropharyngeal cancer patients. Radiother Oncol 2016;120(1):56–62. https://doi.org/10.1016/j.radonc.2016.04.038.
[16] Tattenberg S, Madden TM, Gorissen BL, Bortfeld T, Parodi K, Verburg J. Proton range uncertainty reduction benefits for skull base tumors in terms of normal tissue complication probability (NTCP) and healthy tissue doses. Med Phys 2021;48(9):5356–66. https://doi.org/10.1002/mp.15097.
[17] Wagenaar D, Kierkels RGJ, van der Schaaf A, Meijers A, Scandurra D, Sijtsema NM, et al. Head and neck IMPT probabilistic dose accumulation: Feasibility of a 2 mm setup uncertainty setting. Radiother Oncol 2021;154:45–52. https://doi.org/10.1016/j.radonc.2020.09.001.