Robustness Recipes for Minimax Robust Optimization in Intensity Modulated Proton Therapy for Oropharyngeal Cancer Patients

Van Der Voort, Sebastian; Van De Water, Steven; Perkó, Zoltán; Heijmen, Ben; Lathouwers, Danny; Hoogeman, Mischa

DOI
10.1016/j.ijrobp.2016.02.035

Publication date
2016

Document Version
Accepted author manuscript

Published in
International Journal of Radiation: Oncology - Biology - Physics

*Title Page (WITH Author Details)

Robustness recipes for minimax robust optimization in intensity-modulated proton therapy for oropharyngeal cancer patients

Sebastian van der Voort, BSc[1,2]

Steven van de Water, PhD[1]

Zoltán Perkó, PhD[2,*]

Ben Heijmen, PhD[1]

Danny Lathouwers, PhD[2]

Mischa Hoogeman, PhD[1]

[1]Erasmus MC Cancer Institute, Department of Radiation Oncology, Groene Hilledijk 301, 3075 EA Rotterdam, the Netherlands

[2]Delft University of Technology, Department of Radiation, Science and Technology, Section Nuclear Energy and Radiation Applications, Mekelweg 15, Delft, The Netherlands

*Currently at Massachusetts General Hospital, Department of Radiation Oncology, Physics Research Group, Boston, Massachusetts, US.

Correspondence:

Mischa Hoogeman

Groene Hilledijk 301

3075 EA Rotterdam

The Netherlands

Tel: +31 10 7041748

Fax: +31 10 7041012

m.hoogeman@erasmusmc.nl

Conflict of interest: Erasmus MC Cancer Institute has research collaborations with Elekta AB, Stockholm, Sweden and Accuray Inc, Sunnyvale, USA.

Title: Robustness recipes for minimax robust optimization in intensity-modulated proton therapy for oropharyngeal cancer patients

Running title: Robustness recipes in IMPT

Keywords: IMPT; robust optimization; oropharyngeal cancer, robustness recipe, Polynomial Chaos

Summary

To account for setup and range uncertainties in IMPT, treatment plans can be robustly optimized by including error scenarios in the optimization. For oropharyngeal cancer patients, we derived ‘robustness recipes’ describing the error values of the included scenarios in minimax worst-case robust optimization that provide adequate CTV coverage for given population based distributions of systematic and random setup errors and proton range errors. The application of the recipe resulted in the desired coverage.

Abstract

Purpose: To derive a 'robustness recipe' giving the range (RR) and setup robustness (SR) settings – i.e. the error values - that ensure adequate CTV coverage in oropharyngeal cancer patients for given Gaussian distributions of systematic and random setup errors and range errors (characterized by standard deviations of Σ, σ and ρ respectively) when used in minimax worst-case robust IMPT optimization.

Methods and Materials: For the analysis contoured CT scans of 6 unilateral and 6 bilateral patients were used. An IMPT plan was considered robust if for at least 98% of the simulated fractionated treatments 98% of the CTV volume received ≥ 95% of the prescribed dose. For fast assessment of the CTV coverage for given error distributions (i.e. different values of Σ, σ, and ρ) Polynomial Chaos methods were employed. Separate recipes were derived for the unilateral (ULC) and bilateral (BLC) cases using one patient from each group, and all 12 patients were included in the validation of the recipes.

Results: Treatment plans for bilateral cases are intrinsically more robust than those for unilateral ones. The required RR only depends on the ρ and SR can be fitted by second order polynomials in Σ and σ. The formulas for the derived 'robustness recipes' are: unilateral patients need $SR=-0.15\Sigma^2+0.27\sigma^2+1.85\Sigma-0.06\sigma+1.22$ and RR=3% for ρ =1% and 2%, bilateral patients need $SR=-0.07\Sigma^2+0.19\sigma^2+1.34\Sigma-0.07\sigma+1.17$ and RR=3% and 4% for ρ=1% and 2%, respectively. For the recipe validation 2 plans were generated for each of the 12 patients corresponding to Σ=σ=1.5 mm and ρ=0% and 2%. 22 plans had adequate CTV coverage in ≥98% of the simulated fractionated treatments, the remaining two had adequate coverage in 97.8% and 97.9%.

Conclusions: Robustness recipes were derived that can be used in minimax robust optimization of IMPT treatment plans to ensure adequate CTV coverage for oropharyngeal cancer patients.

Introduction

Intensity-Modulated Proton Therapy (IMPT) can potentially results in improved sparing of Organs At Risk (OARs) compared with Intensity-Modulated Radiotherapy (IMRT) using photon beams [1,2]. However, the accuracy of IMPT dose delivery is highly susceptible to inaccuracies in the patient setup (characterized by a systematic setup error and a random setup error) and the anticipated proton range (known as the range error) [3-6]. In traditional IMRT these uncertainties can be taken into account by margin recipes providing the needed margin around the Clinical Target Volume (CTV) that ensures adequate CTV coverage [7, 8]. Such recipes however cannot be applied to IMPT, because the underlying concept of Planning Target Volume is not applicable to proton therapy [9].

To account for setup and range uncertainties in IMPT, robust treatment plans can be constructed by using 'minimax' optimization [10,11]. In the 'minimax' method the worst-case value of the constraints and the objectives is optimized while simultaneously considering the dose distribution in the nominal scenario (without range or setup errors) and a limited number of error scenarios. Typically these error scenarios correspond to an over- and underestimation of the proton range (i.e. higher and lower range values without setup errors) and rigid shifts of the anatomy along each of the three principle axes (i.e. plus and minus shifts in the x, y and z directions without any range error), resulting in 9 total scenarios (1 nominal, 2 with range and 6 with setup errors). Since the minimax method disregards the probabilities of these scenarios it may lead to overly conservative treatment plans. Conversely, it does not consider the simultaneous occurrence of setup and range errors either and therefore may result in treatment plans that are not sufficiently robust.

Currently it is unknown what the robustness settings – i.e. the absolute values of the errors characterizing the 8 error scenarios - in the 'minimax' robust optimization methods should be in order to obtain a treatment plan that has a specified CTV coverage for a certain fraction of the patients in the presence of setup and range uncertainties. Therefore in this study we derived - for

the first time to our knowledge - 'robustness recipes' that provide the robustness settings - i.e. the error values – that have to be considered during 'minimax' optimization to achieve adequate target coverage in IMPT treatments of unilateral and bilateral oropharyngeal cancer patients. The required error values are referred to as Range Robustness (RR; given in % corresponding to the range error value in the 2 scenarios having only range errors) and Setup Robustness (SR; given in mm corresponding to the magnitude of the shifts along the 3 principal axes in the 6 scenarios with setup errors), and are given for different Gaussian distributions of systematic and random setup errors and proton range errors (characterized by their standard deviations of Σ, σ, and ρ, respectively). To obtain these 'robustness recipes', treatment plans were generated with different robustness settings and their actual robustness was evaluated for different combinations of systematic and random setup error and range error distributions. This required extensive dose calculations, which was made feasible by employing Polynomial Chaos Expansion (PCE) methods that allow for very fast simulations of fractionated treatments [*].

Material and Methods

Patient data

In this study, the CT and contour data of twelve previously treated oropharyngeal cancer patients (six unilateral and six bilateral) was used. Of these twelve patients, one unilateral case (ULC) and one bilateral case (BLC) were selected randomly to be used as training data set to derive the 'robustness recipes' and all 12 patients were used as test data set to validate them. Table 1 shows the patient characteristics.

Recipe derivation methodology

The approach to derive the 'robustness recipes' consisted of four steps. First, the RR was found that was needed to deal with different range errors. Second, treatment plans with different SR were made (in combination with the RR derived in the first step) using XXXX (see section on Treatment planning). The third step was to find for each treatment plan the largest systematic errors Σ that would still give adequate target coverage for different random setup errors σ. Last, the obtained data (i.e. the combinations of Σ and σ for different ρ values) was fit with quadratic polynomials, which together with the data itself form the recipes.

Treatment planning

Treatment plans were generated using XXXX [*], an in-house developed treatment planning system for fully automated, multi-criteria plan generation. Instead of using a single weighted-sum objective function, it optimizes the various defined objectives one-by-one according to priorities assigned by the user in the so-called 'wish-list' (lexicographic optimization). Wish-lists do in general also contain constraints, which always have to be met. Table 2 shows the wish-list used for all patients in this study. The automatic plan generation enabled the inclusion of a large number of treatment plans of consistently high, unbiased quality [*].

To account for setup and range errors XXXX uses the 'minimax' approach as described in the Introduction to make treatment plans robust [10,11 12, 13]. Setup errors were simulated by laterally shifting the pencil beams and range errors were simulated by scaling the values of the CT image.

To select optimal pencil beams, a proton beam resampling technique was used [*]. First, candidate pencil beams were randomly sampled from a very fine grid. Then the multi-criteria optimization was performed and at the end of the iteration the pencil beams with a low contribution were excluded from the next iteration. These three steps were repeated for each optimization iteration.

A dose of 66 Gy was prescribed to the high-dose CTV (primary tumor and positive neck levels) and a dose of 54 Gy to the low-dose CTV (elective neck levels), to be delivered simultaneously in 30 fractions [14]. CTV dose prescriptions were set to obtain for both CTVs at least 98% coverage with at least 95% of the prescribed dose ($V_{95\%} \geq 98\%$) in all 9 scenarios and to keep the CTV volume receiving more than 107% of the prescribed dose below 2% ($V_{107\%} \leq 2\%$). Proton energies that could be used ranged from 70 to 230 MeV and corresponding pencil beam widths ranged from 7 to 3 mm sigma (in air at the isocenter), respectively. A range shifter of 75 mm was used for superficial target regions, where it was assumed that this shifter could be inserted during the delivery of a treatment field. Three beam directions with angles of 60, 180 and 300 degrees were used.

Treatment simulations

To quantify the robustness of a treatment plan we considered the $D_{98\%}$ of the CTV-high and CTV-low for different combinations of systematic and random setup error and range error distributions. The errors were all assumed to have Gaussian distribution with zero mean and standard deviations of Σ, σ and ρ, respectively. E.g. σ =2 mm means a random setup error with a Gaussian distribution with 2 mm standard deviation, while ρ= 2% means a Gaussian range error with a standard deviation of 2%.

To simulate the effects of systematic and random setup errors, as well as range errors on the dose distribution of a fractionated treatment (assuming an infinite number of fractions), we used Polynomial Chaos Expansions (PCEs) [*]. PCEs are meta-models of the dose engine, which can be used to quickly calculate the expected dose distribution of a fractionated treatment for a given distribution of random setup errors and range errors for a specific systematic setup error. This method explicitly takes into account that proton dose distributions are not invariant under setup and range errors. For a detailed description of PCEs and validation experiments for its use in IMPT in oropharyngeal cancer patients we refer to XXXX [*] and the supplementary material. We simulated dose distributions of 100,000 fractionated treatments with different systematic setup and range errors for each combination of Σ, σ and ρ in order to obtain the distribution of the $D_{98\%}$ of the CTV-high and CTV-low.

Study design

A treatment was considered robust (had an adequate CTV coverage) when for 98% of the simulated fractionated treatments (i.e. 98% of the patients) $D_{98\%} \geq 95\%$ of the prescribed dose held for both the CTV-high and the CTV-low. To establish the robustness settings for an adequately robust treatment we first determined the RR that is needed to handle a ρ of 1% or 2% without any setup error (for ρ=0% it was assumed that no RR was required). To this end, treatment plans with RR settings from 0% to 4%, in steps of 1%, were evaluated using the PCE-based simulations described above for ρ= 1% or 2% and Σ= σ= 0mm.

Next, plans were generated with the obtained RR settings combined with SR of 2, 3, 4, or 5mm. Subsequently, we determined the combinations of systematic and random setup errors that still gave an adequate CTV coverage. For this evaluation seven random errors from σ = 0 to 3 mm in steps of 0.5 mm were tested. For each of them we determined the maximum Σ that would still result

in an adequate CTV coverage. Figure 1 shows a flowchart of the method. We first set an initial guess for Σ, then simulated fractionated treatments for the given Σ, σ and ρ, and finally determined the fraction of the treatments which had a $D_{98\%}$ of 95% of the prescribed dose. If this fraction was larger than 98% for both the CTV-high and CTV-low, Σ was increased, if it was less than 98% for either the CTV-high or CTV-low, or both, Σ was decreased. The minimum step size in Σ was 0.05 mm. At some point the previous systematic error did pass the coverage criterion, but the increased error did not. In that case the previous systematic error was taken as the maximum systematic error that could be handled. This was done separately for the chosen ULC and BLC. It is also possible to construct the robustness recipe by searching for the robustness needed to handle a certain error instead of searching for the error that can be handled by a certain robustness. However, it is much more time consuming to construct a treatment plan than constructing a PCE for different errors and do the simulations, therefore the latter approach was chosen.

The obtained robustness settings were verified considering a specific combination of errors for all 12 patients in the test data set. For each patient we used Σ=σ=1.5 mm and ρ=0% or 2% [3]. The Σ and σ of 1.5mm correspond to residual setup errors due to bony anatomy deformation, in case of a daily alignment on vertebrae C1-C3 [15]. The SR settings were calculated by interpolation between the SR settings obtained earlier. For all patients, treatment plans were generated using the interpolated SR settings and the needed RR setting. Next, we simulated 100.000 fractionated treatments for each treatment plan and determined the fraction that received a $D_{98\%}$ of 95%.

To facilitate the usage of the recipe for systematic and random setup errors for which no data points are available, a least squares fit of the form of $SR = a\Sigma^2 + b\sigma^2 + c\Sigma + d\sigma + e$ was done. Since the data suggested (see below) that the SR and RR were practically independent, the data over all ρ values was combined for this fit, thus the range error was not accounted for explicitly.

Results

We found that a 3% RR was needed for ρ = 1% and 2% to achieve an adequate CTV coverage for the ULC. For the BLC, a 3% RR was needed for ρ=1%, whereas a 4% RR was needed for ρ=2% (Table 3).

Figure 2 shows the robustness recipe for different range errors. The data points indicate the maximum combination of systematic and random setup errors for which a certain SR provided adequate CTV coverage. These plots can be used to determine the SR needed, if the error distributions are known. For example, if we assume σ=1.5mm and Σ=1.3 mm and ρ=2%, a 4 mm SR is needed to achieve adequate coverage for the ULC.

Comparing the plots for different range errors, we see that changing the range error (and adjusting the RR accordingly) has a limited impact on the curves for the required SR. Figure 2 also shows that for the BLC the same SR can handle larger setup errors compared with the ULC. This suggests that the treatment plan of the BLC is intrinsically more robust against setup errors.

For the ULC, the fit to the data points in Figure 2 resulted in $SR=-0.15\Sigma^2+0.27\sigma^2+1.85\Sigma-0.06\sigma+1.22$. For the BLC the fit is $SR=-0.07\Sigma^2+0.19\sigma^2+1.34\Sigma-0.07\sigma+1.17$. In Figure 2 these fits are displayed as dashed lines. For combinations of Σ and σ where the recipe in Figure 2 does not have data points the fits can be used, which completes the recipe.

To validate the recipe for all twelve patients for Σ=σ=1.5 mm and ρ=0% or 2%, the required SR was interpolated from the data in Figure 2. The interpolated SR was 4.1 or 4.3 mm (unilateral patients) and 3.3 or 3.4 mm (bilateral patients), for a ρ=0% or 2% respectively. The RR used was 3% for ρ=1% (both unilateral and bilateral patients) and 3% (unilateral patients) or 4% (bilateral patients) for ρ=2%. Table 4 lists for each patient separately the results of the validation in terms of the fraction of

the simulated treatments receiving a $D_{98\%} \geq 95\%$. In two cases (training data set itself) the fraction was below, but close to 98%, i.e. 97.8%. For the other patients, this fraction was above 99%.

Discussion

To the best of our knowledge this is the first study establishing a recipe for calculating the error values for the scenarios to be included in minimax robust optimization that ensure an adequate CTV coverage for fractionated IMPT treatments in the presence of prior known systematic and random setup error and range error distributions. The recipe provides a practical method to bridge the gap between a straightforward minimax worst-case robust optimization and a probabilistic optimization. We found a difference in robustness for unilateral and bilateral patients, where a treatment plan for a bilateral patient seemed to be more robust than a treatment plan for a unilateral patient with similar robustness settings. This suggests, that robustness settings need to be separately tuned for the various treatment sites and groups.

In this study we deemed a treatment to be adequate if in at least 98% of the simulated fractionated treatments the $D_{98\%} \geq 95\%$ of the CTV. In the past, other definitions were used for photon beam therapy. For example, van Herk et al. [7] considered a cumulative minimum CTV dose of 95% of the prescribed dose in 90% of the patients as adequate. In this study we chose the near minimum dose ($D_{98\%}$) following ICRU Report 83 [16]. Van Herk et al. [7] also used maximally tight synthetic dose distributions for the analyses. In this study we used realistic distributions, therefore it was decided to require adequate CTV coverage for 98% of treatments, instead of the 90% used by van Herk.

A limitation of this study was that both recipes were derived based on the CT and treatment plan data of a single patient from each group. In the validation experiments those patients appeared to be the 'worst cases' in their respective groups. For the other patients the CTV coverage was consistently higher than the required 98%. No particular reason was found why the training patients were the most sensitive to setup and range errors. It might be interesting to have a larger training data set. This would likely decrease the needed SR for the same setup errors. However this will most

likely also result in an inadequate tumor coverage for some patients. Future research will focus on gaining insight into inter-patient differences of the recipe and means of individualizing the recipe.

Another limitation of this study is the fact that the required SR and RR were determined independently, potentially resulting in non-optimal robustness recipes. Nevertheless, the required SR values for treatment plans with and without accounting for range errors were not substantially different (Figure 2), which seems to suggest that setup and range robustness can indeed be handled independently. However, the required RR was determined in steps of 1%, therefore the recommended RR settings slightly overcompensate the range errors for which they were obtained (Table 3). The extra robustness may have compensated for interdependencies between setup and range errors.

The difference between the unilateral and bilateral patients seems consistent across patients, as the validation showed that a lower SR for the bilateral patients gives approximately the same CTV coverage as a higher SR for the unilateral patients (Table 4). The mean fraction of the population with a $D_{98\%}$ of 95% for a 0% range error was 99.6% and 99.5% for the unilateral and bilateral patients respectively, not including the patients from the training data set. The bilateral patients, however, needed a higher range robustness setting and even with this higher setting a lower fraction of the simulated fractionated treatments received an adequate CTV coverage compared with the unilateral patients when considering only range errors (Table 3).

The robustness recipes contain a constant term, which means that even if a setup and range error are absent, the recipe indicates that some robustness is still needed. Not including this constant term in the fit drastically lowered the quality of the fit. The smallest error that has been simulated when constructing the robustness recipe was a 0.5 mm systematic setup error. We discourage to use the recipe below this value (performance for error combinations outside the SR=5 mm lines in Figure

2 is not guaranteed either) The robustness recipe was not designed to compensate for anatomical changes, which should be accounted for by re-planning [3].

Conclusions

In this paper we developed for unilateral and bilateral oropharyngeal patients recipes for the error values that should be taken into account in minimax worst-case robust IMPT optimization to ensure robust CTV coverage in case of uncertainties in calculated proton ranges and systematic and random patient setup errors. The recipes for the patient groups were substantially different, as plans for bilateral patients were inherently more robust. A validation study showed that the recipes did indeed result in adequately robust plans.

References

[*] Reference blinded for review

1. van de Water TA, Lomax AJ, Bijl HP, et al. Potential benefits of scanned intensity-modulated proton therapy versus advanced photon therapy with regard to sparing of the salivary glands in oropharyngeal cancer. International Journal of Radiation Oncology Biology Physics 2011; 79(4):1216–1224.
2. Simone CB, Ly D, Dan TD, et al. Comparison of intensity-modulated radiotherapy, adaptive radiotherapy, proton radiotherapy, and adaptive proton radiotherapy for treatment of locally advanced head and neck cancer. Radiotherapy and Oncology 2011;101(3):376–382.
3. Kraan AC, van de Water S, Teguh DN, et al. Dose uncertainties in IMPT for oropharyngeal cancer in the presence of anatomical, range, and setup errors. International Journal of Radiation Oncology Biology Physics 2013;87(5):888–896.
4. Lomax A. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties. Physics in medicine and biology 2008; 53(4):1027–1042.
5. Lomax A. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions. Physics in medicine and biology 2008;53(4):1043–1056.
6. Jiang H, Seco J, Paganetti H. Effects of Hounsfield number conversion on CT based proton Monte Carlo dose calculations. Medical physics 2007;34(4):1439–1449.
7. van Herk M, Remeijer P, Rasch C, et al. The probability of correct target dosage: dose population histograms for deriving treatment margins in radiotherapy. International Journal of Radiation Oncology Biology Physics 2000;47(4):1121–1135.
8. van Herk M, Remeijer P, Lebesque JV. Inclusion of geometric uncertainties in treatment plan evaluation. International Journal of Radiation Oncology Biology Physics 2002;52(5):1407–1422.

9. Liu W, Frank SJ, Li X, et al. PTV-based IMPT optimization incorporating planning risk volumes vs robust optimization. Medical physics 2013;40(2):021709–1–021709–8.
10. Fredriksson A, Forsgren A, Hårdemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. Medical physics 2011;38(3):1672–1684.
11. Chu M, Zinchenko Y, Henderson SG, et al. Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty. Physics in Medicine and Biology 2005;50(23):5463–5478.
12. Liu W, Zhang X, Li Y, et al. Robust optimization of intensity modulated proton therapy. Medical physics 2012;39(2):1079–1091.
13. Chen W, Unkelbach J, Trofimov A, et al. Including robustness in multi-criteria optimization for intensity-modulated proton therapy. Physics in medicine and biology 2012;57(3):591.
14. Widesott L, Pierelli A, Fiorino C, et al. Intensity-modulated proton therapy versus helical tomotherapy in nasopharynx cancer: planning comparison and NTCP evaluation. International Journal of Radiation Oncology Biology Physics 2008;72(2):589–596.
15. van Kranen, S., van Beek, S., Rasch, C. et al. Setup uncertainties of anatomical sub-regions in head-and-neck cancer patients after offline CBCT guidance. International Journal of Radiation Oncology Biology Physics 2009;73(5):1566-1573.
16. Report 83: Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT). Journal of the ICRU 2010;10 (1): 93-106.

Figure legends

Figure 1: Flowchart of the method to determine the maximum systematic error Σ that can be handled by a specific robustness setting to still give an acceptable CTV coverage.

Figure 2: Combinations of random (σ) and systematic (Σ) setup errors that give a $D_{98\%}$ of 95% of the prescribed dose to the CTV for 98% of the fractionated treatments for range errors of 0%, 1% and 2% for a unilateral and bilateral patient. In each plot different SR and RR settings are shown. The solid round markers show the obtained data, the dashed lines are a quadratic fit.

Table

Table 1: Summary of patient characteristics

Patient	Site	Stage	CTV-high (mL)	CTV-low (mL)
1	Soft palate	T2N0	14	72
2	Base of tongue	T1N2c	106	199
3	Base of tongue	T3N2a	68	221
4	Tonsil	T2N0	5	67
5	Tonsil	T1N1	41	95
6	Base of tongue	T3N2a	99	313
7	Tonsil	T2N1	43	165
8	Tonsil	T2N0	11	77
9	Base of tongue	T3N3	178	343
10	Base of tongue	T1N2c	70	294
11	Tonsil	T1N2b	45	279
12	Epiglottic vallecula	T2N0	89	138

Table 2: Planning constraints and objectives applied in this study. The order in which the objectives were optimized is indicated by the priorities where a lower number means a higher priority. The CTV intermediate is a 10 mm transition region between the high-dose and low-dose CTV. The CTV-low' consists of the low-dose CTV excluding the transition region. The robust column indicates whether or not that constraint was robustly optimized.

Constraints				
	Structure	**Type**	**Limit**	**Robust**
	CTV high	Minimum	0.98 ·66Gy	Yes
	CTV intermediate	Minimum	0.98 · 54Gy	Yes
	CTV low	Minimum	0.98 · 54Gy	Yes
Objectives				
Priority	**Structure**	**Type**	**Goal**	**Robust**
1	CTV high	Maximum	1.07 · 66Gy	Yes
1	CTV intermediate	Maximum	1.07 · 66Gy	Yes
1	CTV low'	Maximum	1.07 · 54Gy	Yes
2	CTV high rings(high-dose conformality)	Maximum	1.07 · 66Gy	No
2	CTV combined rings (high-dose conformality)	Maximum	1.07 · 54Gy/ 0.9 ·54Gy	No
3	Parotid	Mean	0Gy	Yes
4	Submandibular glands	Mean	0Gy	Yes
5	Cord	Maximum	20Gy	Yes
5	Brain stem	Maximum	20Gy	Yes
6	Larynx	Mean	0Gy	Yes
6	Oral cavity	Mean	0Gy	Yes
7	Swallowing muscles	Mean	0Gy	Yes
8	CTV rings(low-dose conformality)	Maximum	0Gy	No
8	CTV rings(low-dose conformality)	Mean	0Gy	No
9	Total spot weight	Sum	0Giga-protons	No

Table 3: Percentages of simulated fractionated treatments that passed the $D_{98\%} \geq 95\%$ criterion for different range robustness settings and evaluated range errors, ρ. No setup robustness was used and setup errors were not simulated.

	ρ = 1%		ρ = 2%	
Patient	**RR[1](%)**	**SF[2](%)**	**RR(%)**	**SF(%)**
ULC[3]	1	87.5	1	79.0
	2	92.8	2	85.6
	3	100.0	3	99.8
	4	100.0	4	100.0
BLC[4]	1	0.0	1	0.0
	2	50.5	2	46.9
	3	99.5	3	96.8
	4	99.9	4	98.6

[1] Range robustness

[2] Fraction of simulated treatments passing the coverage criteria

[3] Unilateral case (patient with unilateral cancer)

[4] Bilateral case (patient with bilateral cancer)

Table 4: Validation of generated data points. Presented are percentages of simulated fractionated treatments that passed the $D_{98\%} \geq 95\%$ criterion for a 1.5 mm systematic (Σ) and random setup error (σ) and ρ = 0% and 2%. The robustness settings were derived from the data obtained for patient 1 and 2.

		Fraction of simulated treatments with $D_{98\%} \geq 95\%$ (%)	
Patient	**Treatment group**	**ρ = 0%**	**ρ = 2%**
1	Unilateral	97.9	98.1
2	Bilateral	98.2	97.8
3	Bilateral	99.4	99.1
4	Unilateral	99.3	99.0
5	Unilateral	99.9	100.0
6	Bilateral	99.6	99.6
7	Bilateral	99.5	99.4
8	Unilateral	99.0	99.2
9	Bilateral	99.4	99.5
10	Bilateral	99.7	99.7
11	Unilateral	99.9	99.6
12	Unilateral	100.0	99.9
Average	Unilateral	99.3	99.3
Average	Bilateral	99.3	99.2

Figure1

Click here to download Figure: figure1.pdf

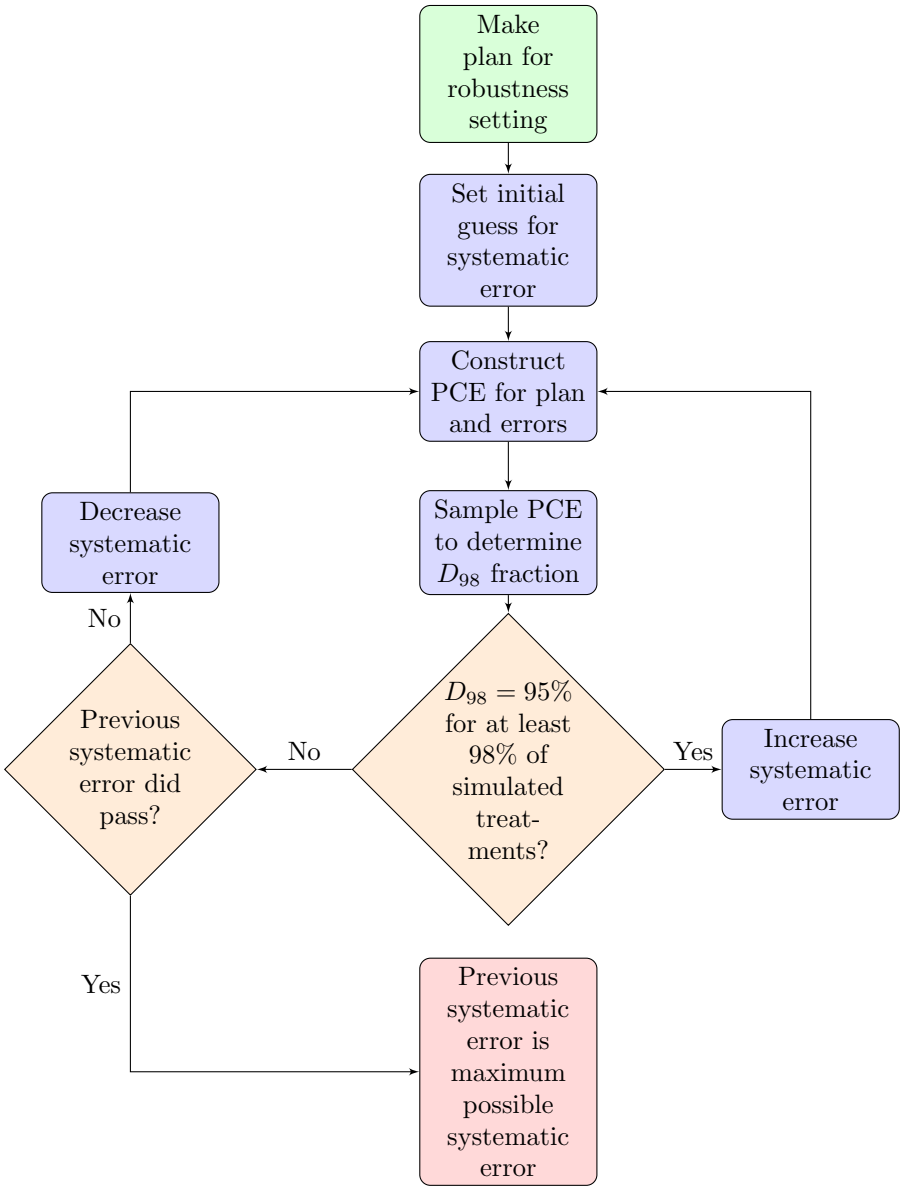

Figure2

Click here to download Figure: figure2.pdf