Received:
02 July 2019

Revised:
10 January 2020

Accepted:
21 January 2020

<https://doi.org/10.1259/bjr.20190584>

Cite this article as:

Teoh S, George B, Fiorini F, Vallis KA, Van den Heuvel F. Assessment of robustness against setup uncertainties using probabilistic scenarios in lung cancer: a comparison of proton with photon therapy. *Br J Radiol* 2020; **93**: 20190584.

PROTON THERAPY SPECIAL FEATURE: FULL PAPER

Assessment of robustness against setup uncertainties using probabilistic scenarios in lung cancer: a comparison of proton with photon therapy

^{1,2}SULIANA TEOH, FRCR, ^{1,2}BEN GEORGE, PhD, ^{1,3}FRANCESCA FIORINI, PhD, ^{1,2}KATHERINE A VALLIS, PhD, FRCR and ^{1,2}FRANK VAN DEN HEUVEL, PhD, FAPM

¹CRUK and MRC Oxford Institute for Radiation Oncology, Old Road Campus Research Building, Off Roosevelt Drive, University of Oxford, Oxford, OX3 7DQ, UK

²Department of Radiotherapy, Oxford Cancer Centre, Oxford University Hospitals NHS Foundation Trust, Oxford OX3 7LE, UK

³Rutherford Cancer Centre Thames Valley, 3 Collegiate Square, Shinfield, Reading RG2 9LH, UK

Address correspondence to: Suliana Teoh

E-mail: suliana.teoh@oncology.ox.ac.uk

Objective: We compared the sensitivity of intensity modulated proton therapy (IMPT) and photon volumetric modulated arc therapy (VMAT) plans to setup uncertainties in locally advanced non-small cell lung cancer (NSCLC) using probabilistic scenarios.

Methods: Minimax robust (MM) and planning target volume (PTV) optimised IMPT and VMAT nominal plans were created with physical dose of 70 Gy in 35 fractions in 10 representative patients. Using population data of setup errors, a fractionated treatment course was simulated, summed (D_{sum}) and compared to the nominal plan. Three treatment-course simulations were done for each plan. Target robustness criteria were: dose deviation of $\leq 5\%$ to clinical target volume (CTV) $D_{98\%}$ and CTV $V_{95\%} \geq 99.9\%$. Voxelwise simulation repeatability was analysed using Bland–Altman plots. Acceptable limits of agreement were 2% of the prescription dose.

Results: All D_{sum} met target robustness criteria. While fraction VMAT and MM-IMPT doses were excellent,

simulated fraction doses in PTV-IMPT were suboptimal. Almost all ($>99\%$) of VMAT and MM-IMPT fraction doses met both target robustness criteria. For PTV-IMPT, only 96.9 and 80.3% of fractions met $CTVD_{98\%}$ and $V_{95\%}$ criteria respectively. Simulation repeatability was excellent (limits of agreement range: 0.41–1.1 Gy) with strong positive correlations.

Conclusion: When considering the whole treatment course, setup errors do not influence robustness irrespective of planning techniques used. However, on a fraction level, VMAT and MM-IMPT plans are superior compared to PTV-IMPT plans.

Advances in knowledge: Probabilistic analysis provides a fast and practical method for evaluating VMAT and IMPT plan sensitivity against setup uncertainty. VMAT and robust-optimised IMPT plans have comparable sensitivity to setup uncertainties in conventionally fractionated treatment for NSCLC.

INTRODUCTION

Proton beam therapy (PBT) holds the promise of improving cure without increasing treatment-related toxicity due to its highly localised dose deposition at the Bragg peak. However, accurate placement of the Bragg peak is necessary if under dosing the target or overdosing organs at risk (OAR) are to be avoided. Sources of uncertainty in PBT include setup, anatomical variations, dose calculation algorithms, organ motions and biological effects.¹ The sensitivity of a plan to various uncertainties is termed *robustness*.

In photon therapy, setup errors are compensated for by adding a margin to the treatment volume to produce a planning target volume (PTV).² This may be inadequate in

PBT as the margin required is highly dependent on tissue density changes along the beam path. A standard PTV does not account for the non-uniform distortion of the proton dose distribution caused by a setup error. This situation is particularly concerning when non-uniform fields—multi-field optimisation (MFO), are used. Incorporating uncertainties in MFO plan is not straightforward. Margins can improve target coverage only at the edges of the target volume and not within the target itself and have little effect on the sensitivity of the plan when there are steep dose gradients within the clinical target volume (CTV).³

Alternatively, setup and range uncertainties can be accounted for using robust optimisation.^{4–7} This approach

does not require a PTV to be created. Current commercially available treatment planning systems (TPS) use minimax optimisation which optimises the worst-case scenario dose distribution at each iteration.⁶

Following optimisation, evaluation of plan sensitivity is necessary. Several methods have been reported for quantifying plan robustness.^{3,8,9} There is currently no consensus on the definition of robustness or on the ideal metric for assessment. Some methods rely on simulating treatment plans under a number of extreme conditions: worst-case dose distribution,⁸ error-volume histogram³ and dose-volume histogram (DVH) bands,⁹ which can overestimate errors. Furthermore, the assumption that larger shifts result in worse errors than smaller ones may not be true,¹⁰ especially in the presence of high tissue heterogeneity (*i.e.* thoracic intensity modulated proton therapy (IMPT)). Other analyses use statistical sampling where several hundred scenarios are simulated.^{11,12} These may be more realistic, but may not be practical for routine clinical use due to the time it takes to simulate the scenarios.

In the treatment of moving targets, plans need to be robust against various uncertainties. We describe a practical step in assessing one of these uncertainties—setup uncertainty in non-small cell lung cancer (NSCLC) using probabilistic scenarios. Using a simple extension of a commercial TPS, a fractionated treatment course was simulated, summed and compared to the nominal treatment plan. Our initial results appeared promising with robust-optimised IMPT plans having similar target robustness to volumetric modulated arc therapy (VMAT).¹³ The time taken to simulate a 35-fraction treatment course was approximately 5 minutes on a Xeon-based workstation. The method was independent of the TPS optimisation algorithm, thereby providing an independent measure of robustness. The main requirement for the probabilistic method is knowledge of the population setup errors, which can be derived by individual institutions. This is advantageous for new proton centres or treatment of new tumour sites where retrospective robustness analysis of PBT treatments is not possible.¹⁴ If IMPT is to be a

viable alternative, plans created need to be robust and its robustness needs to be comparable to photon therapy. Therefore, we compared the robustness of IMPT to photon VMAT.

METHODS AND MATERIALS

Plans

10 patients with NSCLC treated with photon radiotherapy between August 2013 and March 2017 at the Oxford Cancer Centre were selected based on varying tumour sizes and locations (Table 1). The planning study was registered and approved by the NHS Health Research Agency and conducted under the auspices of Oxford University Clinical Trials and Research Governance (research ethics committee reference: 16/LO/1324).

Four-dimensional (4D) CT simulation (Optima CT580RT, GE Healthcare, WI) were acquired for all patients. Each 4DCT data set was acquired under free breathing conditions and recorded using the Real-Time Position Management system (Varian Medical Systems, Palo Alto, CA). The entire thorax and upper abdomen were scanned with 2.5 mm slice thickness. Each breathing period was phase-binned into 10 bins. Unweighted averaged-intensity projection (Ave-CT) data sets were generated for treatment planning.

Contouring

Target and normal anatomy delineation were done by a Radiation Oncologist (ST) following the RTOG1308 trial protocol.¹⁵ To account for tumour motion, the gross tumour volume (GTV) was contoured in the maximum and minimum inhalation phases and combined to form the internal target volume (ITV). The ITV was verified in all 4DCT phases. An 8 mm isotropic expansion of the ITV, edited to anatomical boundaries, formed the CTV. The CTV-PTV margins were calculated using $2.5\Sigma + 0.7\sigma$, where Σ is the population setup mean systematic error and σ is the corresponding population mean random error.¹⁶ Oxford Cancer Centre mean population systematic errors, Σ , were 1.0, 1.3, and 0.8 mm and random errors, σ , were 2.3, 2.7, and 2.3 mm in the x-, y- and z- directions respectively. This is calculated from recorded

Table 1. Tumour characteristics

Patient	GTV (cc)	TNM seven staging	Primary tumour location	IASLC lymph node stations
1	15	TxN2M0	-	5
2	25	TxN2M0	-	10R, 4R
3	27	T2N1M0, T1N0M0	Left lower lobe and upper lobe	10L
4	28	T1N2M0	Left upper lobe	10L, 4L
5	32	T2N0M0	Right lower lobe	-
6	50	T2N2M0	Right upper lobe	4R
7	106	T2N0M0	Right middle lobe	-
8	115	T4N1M0	Right lower lobe	10R
9	261	T4N2M0	Left upper lobe	7, 10L
10	404	T4N0M0	Left upper lobe	-

GTV, gross tumour volume.

Tx – undetectable primary.

daily online shifts of patients who underwent radical radiotherapy for lung cancer at our institution. The maximum calculated distance defined the CTV-PTV isotropic margin (5 mm).

Treatment planning

For every patient clinical nominal plans (D_{nom}) were generated for each optimisation technique: VMAT, PTV-optimised IMPT (PTV-IMPT) and two MFO robustly optimised IMPT (MM-IMPT). VMAT plans were designed with two 360° arcs using 6 MV beams.

Unlike photon plans, IMPT plans need to account for both setup and range errors. The robust-optimiser expand the CTV for both of these errors to varying degrees. Therefore for fair comparison, MM-IMPT plans optimised with only setup parameter (MM-IMPT(S)) were created. The second plan was optimised with both setup and range uncertainties parameters (MM-IMPT(S + R)). IMPT plans used three to four beams. The same beam angles were used for both PTV- and MM-IMPT plans (see [Appendix A](#) for beam arrangements of all the plans). As far as possible, use of range shifters were minimised due to limitations of the analytical dose engine.¹⁷ Beam arrangements chosen were based on least anatomical distance to the tumour to reduce irradiation of normal tissue. Use of anterior beams were also minimised to reduce the error resulting from breathing motion. Where tumour motion is not more than 7 mm our analysis have shown that use of the minimax robust optimisation is sufficiently robust against tumour motion (data not shown, unpublished work).

The beam data used was based on an IBA cyclotron at Provision Proton Therapy Centre, Knoxville, TN (98–230 MeV).¹⁸ The dose calculation grid resolution was $2 \times 2 \times 2 \text{ mm}^3$. All treatment planning and dose calculations were performed on Raystation 5.0.2 TPS (RaySearch Laboratories, Stockholm).

Dose specifications and constraints

The target prescription was 70 Gy (relative biological effectiveness (RBE)) in 35 fractions. RTOG-1308 trial dose specifications and constraints were used.¹⁵ Proton RBE was assumed to be 1.1. In the PTV-optimised plans, plans were normalised such that the prescribed dose covered 95% of the PTV. In the CTV-optimised plans, plans were normalised such that the prescribed dose covered 99% of the CTV. Plans were normalised such that 70 Gy (RBE) the prescribed dose covered 95% of the PTV and 70 Gy (RBE) covered or 99% of the CTV in MM-IMPT plans. In all treatment modalities, the prescription dose had to cover at least 99% of CTV.

Robust optimisation parameters

A minimax robust optimisation method was used for MM-IMPT.⁶ The minimax optimisation minimised the objective function in the worst-case scenario. An estimate of the effects of the errors was obtained by recalculating a set of dose distributions based on the scenarios. The TPS would then expand the CTV based on this. The optimisation parameters for setup error was 3 mm based on the institution's threshold for online shifts and 3.5% for range uncertainty as commonly done in

multiple proton centres.¹ The range uncertainty component is attributed to CT imaging and conversion to stopping power which is considered to be a systemic error with a constant value for a specific patient throughout the treatment course.¹

Robustness evaluation against setup uncertainty

For every plan, a set of 35 shifts, representing a treatment course of 35 fractions, in the x, y and z directions were generated. A fraction dose is calculated by shifting the isocentre in the x, y and z direction. In total 30 sets of shifts were generated (3 simulations for 10 patients). The maximum shift was limited to $\pm 3 \text{ mm}$ for each of the three Cartesian planes. The shift was limited to this maximum distance due to the protocol in place at our institution for repeat imaging during radiotherapy setup.

For each plane, shifts were generated as follows:

Using the calculated mean and standard deviation of the population systematic and random errors, we first defined the *patient-specific systematic error* by sampling from a normal distribution, denoted as follows:

$$\Sigma_s \in N(\mu_s, \sigma_s^2) \quad (\text{A.1})$$

where μ_s is the institution's population mean systematic error and σ_s is the standard deviation of the systematic error. Secondly, we defined the *patient-specific random error* by sampling from a normal distribution:

$$\Sigma_r \in N(\mu_r, \sigma_r^2) \quad (\text{A.2})$$

where μ_r is the institution's population mean random error and σ_r is standard deviation of the random error.

Finally, for each simulation, 35 shifts were obtained by sampling from a normal distribution with mean, Σ_s , and standard deviation, Σ_r , determined as follows:

$$\text{Shifts} \in N(\Sigma_s, \Sigma_r) \quad (\text{B})$$

The arithmetic sum of the voxel doses from each of the fractions represent a treatment course (D_{sum}). Three treatment-course simulations were performed for each plan ($35 \times 4 = 140$ dose calculations). The simulations were performed in RayStation using the scripting application programming interface accessed with IronPython.

Robustness evaluations

Criteria

Each D_{sum} was compared to its respective nominal plan. In all cases, OAR dose constraints should be met. Target robustness criteria were:

- Dose deviation of $\leq 5\%$ to CTV $D_{98\%}$ as recommended by the Particle Therapy Co-operative Group consensus guidelines.¹⁹
- CTV $V_{95\%} \geq 99.9\%$ to mirror the target coverage criterion from which the PTV margin was derived.²⁰

Fraction dose analysis

Each fraction dose was evaluated against the target robustness criteria and OAR constraints.

Simulation repeatability

We evaluated the reliability of a single simulation in predicting target robustness by comparing the voxel doses within the CTV of D_{sum} using Bland–Altman plots²¹ and Pearson correlation coefficient. A single simulation would be considered representative of many if the difference in dose for a given voxel between simulations were no more than 1.4 Gy (limits of agreement (LoA) = 2% of the prescription dose).

Analysis

Conformity indices (CI = 95% isodose volume/CTV volume) were calculated. CI and dose metrics were compared using Wilcoxon signed-rank test. The variance of the difference in voxel doses between D_{nom} and D_{sum} were calculated for the three treatment modalities. In all the statistical analyses, a p -value of < 0.05 was considered to be statistically significant. All statistical analyses were done in Matlab 2015b, The Mathworks, Natick, 2015.

RESULTS

Nominal plans

CI for both MM-IMPT was statistically significantly lower compared to that for PTV-IMPT ($p < 0.05$). Dose to the heart

(mean heart dose, V5, V30), lung (MLD, V5 and V20) and spinal cord (Dmax) was significantly lower with IMPT compared to VMAT ($p < 0.05$).

Summed and fraction dose distributions

Three treatment course simulations consisting of 35 fractions each were simulated in all 10 patients (4200 fraction doses = 3 simulations x 35 fractions x 4 plans x 10 patients). All D_{sum} met target robustness criteria. One D_{sum} plan for PTV-IMPT (Patient 9) exceeded the maximum dose constraint for the spinal cord (simulation 3, maximum dose to the spinal cord was 50.2 Gy). All other D_{sum} met OAR constraints.

When fraction doses were examined, plan qualities were variable and particularly so for PTV-IMPT plans. Almost all VMAT and both MM-IMPT plans fraction CTV $D_{98\%}$ doses were within 5% of the respective nominal plans (VMAT-99.9%; MM-IMPT(S) and (S + R)-100%). However, only 96.9% (1017/1050 fractions) of PTV-IMPT fractions were within 5% of the nominal CTV $D_{98\%}$. Similarly, almost all VMAT and MM-IMPT fractions met the $V_{95\%}$ criterion (VMAT-100%; MM-IMPT(S)-99.4%, (S + R)-99.5%) compared to 80.3% (843/1050) for PTV-IMPT (Figure 1a). The proportion of fractions per simulation meeting the minimum dose criteria varied most widely for PTV-IMPT (range 54–100%) as can be seen in the boxplots in Figure 2.

Figure 1. Summary of: (A) target coverage (number indicates percentage out of 35 fractions) and (B) adherence to OAR for fraction doses in the different treatment modalities. OAR, organs at risk.

Figure 2. Boxplots of all fractions CTV $D_{98\%}$ and $V_{95\%}$. PTV-IMPT has the largest spread of fraction doses compared to VMAT and both MM-IMPT plans. CTV, clinical target volume; IMPT, intensity modulated proton therapy; PTV, planning target volume; VMAT, volumetric modulated arc therapy

A proportion of fractions failed to meet OAR constraints: six VMAT, five PTV-IMPT, three MM-IMPT(S) and three (MM-IMPT(S + R) simulated plans. Doses to OAR are generally

lowest with MM-IMPT(S), followed by MM-IMPT(S + R) and PTV-IMPT. Details on fraction dose OAR constraint adherence is summarised in [Figure 1b](#). The lung V5 was exceeded in

Figure 3. Dose volume histograms of Patient 10. Nominal versus summed dose distributions and nominal versus perturbed fraction dose distributions with different the treatment modalities. CTV, clinical target volume; IMPT, intensity modulated proton therapy; MM, minimax robust; PTV, planning target volume; VMAT, volumetric modulated arc therapy.

a proportion of simulated fraction (maximum lung V_5 : 66.4%, range 17–25 fractions) in Patient 3. For Patient 9, where the tumour was immediately adjacent to the spinal cord, spinal cord $D_{\max}$ was exceeded in a number of fractions for all three simulations for VMAT (nominal $D_{\max}$ 46.9 Gy; range, 12–19 fractions), PTV-IMPT (48.2 Gy; 14–19 fractions), MM-IMPT(S) (44.0 Gy; 5–12 fractions) and MM-IMPT(S + R) (45.3 Gy; 10 fractions each).

DVH and coronal dose colour maps of a representative patient (Patient 10) are shown in Figures 3 and 4. Visual inspection of the DVHs shows differences between the nominal and summed dose distributions in all three plans. Figure 3b shows DVHs for fraction D_{nom} vs fraction dose distributions. There is little deviation in target coverage with VMAT and MM-IMPT plans on a single fraction level. In contrast, target coverage is inferior in PTV-IMPT plan with sub volumes receiving over- and under-dosage. However, despite the variation of the areas of under-/overdosage with each fraction the sum of all fractions still met the robustness criteria (see Appendix B). A coronal view of the difference between the nominal and summed dose distributions of the PTV-IMPT plan is shown in Figure 4 (PTV-IMPT C). An area of cold spot of up to 2 Gy was observed (indicated by red arrow). This was not observed in the other treatment modalities.

There was internal consistency in the OAR doses for each simulation. IMPT plans deliver statistically significantly lower spinal cord $D_{\max}$, lung $V_{5\text{Gy}}$, lung $V_{20\text{Gy}}$, MLD, heart $V_{30\text{Gy}}$ and heart

$V_{5\text{Gy}}$ compared to VMAT ($p < 0.005$). There was no statistically significant difference between the techniques with respect to heart $V_{45\text{Gy}}$ and oesophagus $V_{50\text{Gy}}$. A comparison of OAR dose metrics in VMAT and IMPT D_{sum} is shown in Appendix C.

Simulation repeatability

Considering the difference in dose between nominal and simulated plans, VMAT had the smallest variance followed by MM-IMPT, then PTV-IMPT (VMAT - 0.78 Gy²; MM-IMPT(S) - 0.97 Gy², MM-IMPT(S + R) - 1.21 Gy²; PTV-IMPT - 2.52 Gy²).

Voxelwise CTV dose analysis showed strong positive correlations between each simulation for all plans (range of r : 0.89–0.94, $n = 107500$, all $p < 0.001$) with LoA ranging from 0.4 to 1.1 Gy (Appendix D and E). All simulations met *a priori* criteria of LoA < 1.4 Gy meaning, for 95% of the target voxel doses, the difference in dose between each simulation was within 2% of the prescribed dose. Similarly, in all cases the 95% confidence interval of the mean difference, upper and lower limit of LoA were no more than 0.02. VMAT and MM-IMPT had superior agreement compared to PTV-IMPT. Appendix D shows the Pearson correlation statistics between each simulation for the different treatment modalities.

DISCUSSION

The sensitivity of PBT plans to uncertainties is a concern in lung cancer due to tissue heterogeneity. The use of MFO with inhomogeneous dose distribution within the target reduces

Figure 4. Coronal views of dose colour map of Patient 10—VMAT, PTV-IMPT and MM-IMPT(S) and MM-IMPT(S + R) plans. IMPT, intensitymodulated proton therapy; MM, minimax robust; PTV, planning target volume; VMAT, volumetric modulated arc therapy. CTV—red contour, 4A—nominal plan, 4B—summed dose distribution, 4C—difference in dose between nominal plan and sum dose distribution. Red arrow indicates area of cold spot within CTV in PTV-IMPT plan.

VMAT

PTV-IMPT

MM-IMPT (S)

MM-IMPT (S+R)

plan robustness. Solutions used in passive scatter and single-field optimised plans such as beam-specific PTV may not hold in highly modulated fields as margins only ensure coverage at the edge of a target and not within it. Use of margins alone may

result in uncertainties within the target.³ In this analysis, we evaluated and compared the robustness of plans against setup errors. Robust-optimised plans were created with the minimax robust

optimisation algorithm using a method independent of the optimisation algorithm.

We evaluated plan robustness using probabilistic scenarios for setup uncertainty in locally advanced NSCLC. Although, the method of sampling from a normal distribution is not new,^{22,23} the probabilistic method is novel because each patient is assigned a systematic and random error based on the population setup error. By sampling from various normal distributions, we assumed that smaller setup errors were more likely to occur than larger ones. All summed simulated doses met the robustness criteria for the target for all patients in VMAT, PTV-IMPT and both MM-IMPT plans. Similar results have been reported in other studies which showed that fractionation reduced the effect of setup uncertainty.^{24,25} However, evaluation of the fraction doses revealed marked differences in plan quality. VMAT and MM-IMPT fractions had comparable target robustness and were superior to PTV-IMPT plans. For PTV-IMPT plans, the minimum dose coverage criteria to the CTV was only fulfilled in 80% of the simulations.

The minimax algorithm produced robust plan to setup error in locally advanced lung cancer. Little variation was observed in fraction and summed dose distribution in the MM-IMPT(S) plans. In PBT, range uncertainty associated with CT and conversion to stopping power exists.¹ As this error needs to be accounted for during planning, we reported the sensitivity analysis for MM-IMPT plans optimised with setup and range parameter.

Typically, the aim of conventional fractionation is to deliver a homogenous dose distribution, the biological effect of the dose variation to the target in IMPT is not known. Simple dose accumulation of the fractions in 2Gy-equivalent doses (EQD2) showed target robustness against both criteria was maintained in all treatment techniques. Conversely, robustness is likely to be an issue if PTV-optimisation is to be used for hypofractionated treatment as hyperfractionation improves plan robustness by compensating for areas that are under/overdosed.

Where OAR constraints were exceeded for fraction doses, care must be taken when deciding if a nominal plan is acceptable,

even if the summed dose distribution meet the constraints. In such cases, the nominal OAR doses are usually close to the OAR constraints.

Comparison of D_{sum} within the CTV at voxel level showed excellent repeatability with LoA within 2% of the prescription dose and strong positive correlations (within 1% for VMAT and MM-IMPT). Given the large number of individual random numbers in each simulation and the excellent agreement between the results from three simulation, these results suggest that the performing a single simulation is representative of multiple simulations for evaluating robustness of the target dose coverage.

We recognised that multiple uncertainties exist in the treatment of moving targets. Setup uncertainty is one aspect that needs to be considered in the treatment of thoracic tumours. In the presence of high tissue heterogeneity, any systematic or random error could result in large dose perturbations. Therefore, assessment of setup uncertainty is just as important as assessment of other uncertainties such as anatomical changes and interplay effects. Furthermore, when investigating different uncertainties, it is crucial to investigate the different sources individually to identify areas that need to be addressed. Assessment of range uncertainty has not been included in this study. A population study is required to use the probabilistic method to assess the effect of range uncertainty. Robustness evaluation of this optimisation algorithm by scaling to the maximum and minimum expected range error has previously been described.⁶ Finally, we recognised the limitation of the pencil beam algorithm for dose calculations. Better agreement with physical measurements were found with Monte Carlo algorithms.^{18,26} Further work is focused on combining the probabilistic scenarios method with interplay using a commercial Monte Carlo dose engine.

In conclusion, probabilistic analyses provide a practical method for determining the sensitivity of VMAT and IMPT plans to setup uncertainties in patients with NSCLC. Whilst target robustness between VMAT and MM-IMPT fractions were excellent, fraction doses in PTV-IMPT were suboptimal.

REFERENCES

1. Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Phys Med Biol* 2012; **57**: R99–117. doi: <https://doi.org/10.1088/0031-9155/57/11/R99>
2. van Herk M. Errors and margins in radiotherapy. *Semin Radiat Oncol* 2004; **14**: 52–64. doi: <https://doi.org/10.1053/j.semradonc.2003.10.003>
3. Albertini F, Hug EB, Lomax AJ. Is it necessary to plan with safety margins for actively scanned proton therapy? *Phys Med Biol* 2011; **56**: 4399–413. doi: <https://doi.org/10.1088/0031-9155/56/14/011>
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. doi: <https://doi.org/10.1088/0031-9155/52/10/009>
5. Pflugfelder D, Wilkens JJ, Oelfke U. Worst case optimization: a method to account for uncertainties in the optimization of intensity modulated proton therapy. *Phys Med Biol* 2008; **53**: 1689–700. doi: <https://doi.org/10.1088/0031-9155/53/6/013>
6. Fredriksson A, Forsgren A, Hårdemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. *Med Phys* 2011; **38**: 1672–84. doi: <https://doi.org/10.1118/1.3556559>
7. Liu W, Zhang X, Li Y, Mohan R. Robust optimization of intensity modulated proton therapy. *Med Phys* 2012; **39**: 1079–91. doi: <https://doi.org/10.1118/1.3679340>
8. 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. doi: <https://doi.org/10.1088/0031-9155/53/4/015>
9. Trofimov A, Unkelbach J, DeLaney TE, 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. doi: <https://doi.org/10.1016/j.prro.2011.08.001>
 10. Casiraghi M, Albertini F, Lomax AJ. Advantages and limitations of the 'worst case scenario' approach in IMPT treatment planning. *Phys Med Biol* 2013; **58**: 1323–39. doi: <https://doi.org/10.1088/0031-9155/58/5/1323>
 11. 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. doi: <https://doi.org/10.1016/j.ijrobp.2013.04.009>
 12. van der Voort S, van de Water S, Perkó Z, Heijmen B, Lathouwers D, Hoogeman M, et al. Robustness recipes for Minimax robust optimization in intensity modulated proton therapy for oropharyngeal cancer patients. *Int J Radiat Oncol Biol Phys* 2016; **95**: 163–70. doi: <https://doi.org/10.1016/j.ijrobp.2016.02.035>
 13. Teoh S, George B, Fiorini F, Vallis K, Van den Heuvel F. Robustness assessment using probabilistic scenarios of intensity modulated proton therapy and volumetric Arc therapy in non-small-cell lung cancer: an in-silico radiotherapy planning study. *The Lancet* 2017; **389**(S1): S94. doi: [https://doi.org/10.1016/S0140-6736\(17\)30490-7](https://doi.org/10.1016/S0140-6736(17)30490-7)
 14. McGowan SE, Albertini F, Thomas SJ, Lomax AJ. Defining robustness protocols: a method to include and evaluate robustness in clinical plans. *Phys Med Biol* 2015; **60**: 2671–84. doi: <https://doi.org/10.1088/0031-9155/60/7/2671>
 15. Liao ZB, Choi J.; N.; Simone CB. RTOG 1308 protocol information. 2015;.
 16. The Royal College of radiologists, society and College of Radiographers, Institute of physics and engineering in medicine. on target: ensuring geometric accuracy in radiotherapy. *The Royal College of Radiologists* 2008;.
 17. Teoh S, Fiorini F, George B, Vallis KA, Van den Heuvel F. Is an analytical dose engine sufficient for IMPT in lung cancer? *British Journal of Radiology* 2019; doi: <https://doi.org/10.1259/bjr.20190583>
 18. Fiorini F, Schreuder N, Van den Heuvel F. Technical note: defining cyclotron-based clinical scanning proton machines in a FLUKA Monte Carlo system. *Med Phys* 2018; **45**: 963–70. doi: <https://doi.org/10.1002/mp.12701>
 19. Chang JY, Zhang X, Knopf A, Li H, Mori S, Dong L, et al. Consensus guidelines for implementing Pencil-Beam scanning proton therapy for thoracic malignancies on behalf of the PTCOG thoracic and lymphoma Subcommittee. *Int J Radiat Oncol Biol Phys* 2017; **99**: 41–50. doi: <https://doi.org/10.1016/j.ijrobp.2017.05.014>
 20. 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. doi: [https://doi.org/10.1016/S0360-3016\(00\)00518-6](https://doi.org/10.1016/S0360-3016(00)00518-6)
 21. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307–10.
 22. Rudat V, Flentje M, Oetzel D, Menke M, Schlegel W, Wannenmacher M. Influence of the positioning error on 3D conformal dose distributions during fractionated radiotherapy. *Radiother Oncol* 1994; **33**: 56–63. doi: [https://doi.org/10.1016/0167-8140\(94\)90086-8](https://doi.org/10.1016/0167-8140(94)90086-8)
 23. Lam KL, Ten Haken RK, Litzenberg D, Balter JM, Pollock SM. An application of Bayesian statistical methods to adaptive radiotherapy. *Phys Med Biol* 2005; **50**: 3849–58. doi: <https://doi.org/10.1088/0031-9155/50/16/013>
 24. Lowe M, Albertini F, Aitkenhead A, Lomax AJ, MacKay RI. Incorporating the effect of fractionation in the evaluation of proton plan robustness to setup errors. *Phys Med Biol* 2016; **61**: 413–29. doi: <https://doi.org/10.1088/0031-9155/61/1/413>
 25. Wölfelschneider J, Friedrich T, Luchtenborg R, Zink K, Scholz M, Dong L, et al. Impact of fractionation and number of fields on dose homogeneity for intra-fractionally moving lung tumors using scanned carbon ion treatment. *Radiother Oncol* 2016; **118**: 498–503. doi: <https://doi.org/10.1016/j.radonc.2015.12.011>
 26. Taylor PA, Kry SF, Followill DS. Pencil Beam Algorithms Are Unsuitable for Proton Dose Calculations in Lung. *Int J Radiat Oncol Biol Phys* 2017; **99**: 750–6. doi: <https://doi.org/10.1016/j.ijrobp.2017.06.003>