Balancing Robustness and Adaptation Rate for Proton Therapy of Lung Cancer Patients

Vlad Badiu[1,2], Vicki Trier Taasti[3], Gilles Defraene[1], Wouter van Elmpt[3], Edmond Sterpin[1,2,4]

[1] KU Leuven, Department of Oncology, Laboratory of Experimental Radiotherapy, Leuven, Belgium

[2] UZ Leuven, Particle Therapy Interuniversity Center Leuven – PARTICLE, Leuven, Belgium

[3] Maastricht University Medical Centre+, Department of Radiation Oncology (Maastro), GROW School for Oncology and Reproduction, Maastricht, The Netherlands

[4] Université catholique de Louvain, Institut de Recherche Expérimentale et Clinique, Center of Molecular Imaging, Radiotherapy and Oncology (MIRO), Brussels, Belgium

Abstract

Introduction: An increase in plan robustness leads to a higher dose to adjacent organs-at-risk (OARs), and an increased chance of post-treatment toxicities. In contrast, more conformal plans lead to sparing of healthy surrounding tissue at the expense of a higher sensitivity to anatomical changes, requiring costly adaptations. In this study, we assess the trade-off and impact of treatment plan robustness on the adaptation rate.

Method: Treatment planning was performed for 40 lung cancer patients, each having a planning 4DCT and up to eight weekly repeated 4DCTs (reCTs). For each patient, plans were made with three different levels of robustness based on setup uncertainty of 3, 6 and 9 mm. These plans were robustly re-evaluated on all reCTs to assess whether the clinical constraints were met.

Results: For the 3, 6 and 9 mm robustness levels, adaptation rates of 87.5%, 70.0% and 57.5%, respectively, were observed. A mean absolute normal tissue complication probability (NTCP) gain of 2.9 percentage points (pp) was calculated for pneumonitis grade ≥2 when transitioning from 9 mm plans to 3 mm plans, 7.6 pp for esophagitis grade ≥2, and 2.5 pp for mortality risk 2 years post-treatment.

Conclusion: The lowered risk of post treatment toxicities at lower robustness levels is clinically relevant but comes at the expense of more treatment adaptations, particularly in cases where meeting our clinical goals is not compromised by having a dose that is more conformal to the target. The trade-off between workload and reduced NTCP needs to be individually assessed.

Keywords: proton therapy, adaptation, robustness, lung cancer

1. Introduction

Proton therapy aims at accurately delivering high radiation doses to tumors while reducing exposure to surrounding healthy tissue when compared to classical photon-based radiotherapy. The steep dose fall-off at the distal edge of the protons' range (the so-called "Bragg Peak") results in a sharply localized dose peak [1]. This improved conformity when compared to photons comes at the expense of an increased susceptibility to treatment errors such as patient positioning or setup errors, range errors (resulting from the conversion of the CT numbers to proton stopping power ratios (SPRs)), and inter- and intra- fraction motion [2–5]. The latter is particularly relevant for tumors of the thorax due to breathing motion potentially displacing density heterogeneities leading to shifts or distortions of the dose distribution. Changes in anatomy can also occur between treatment fractions such as tumor displacements, tissue deformation, weight loss or tumor shrinkage, all of which can lead to a significantly altered dose distribution. This can be addressed by plan adaptation, which takes advantage of high-quality daily volumetric imaging and state-of-the-art replanning tools to ensure the delivered dose matches the planned dose for the new anatomy [6]. The benefit of systematic adaptation for non-small-cell lung carcinoma (NSCLC) have been documented by Koay et al. concluding that adaptive planning leads to better outcomes in terms of ensuring target coverage and reducing organ-at-risk (OAR) dose, as well as reducing the potential health complications that can occur if plans are not adapted to the patient's anatomical changes [7]. Unfortunately, proton treatment plan adaptation, either online or offline, is work and cost demanding, as it typically requires an almost entirely new treatment planning procedure [8]. It should be noted that more time-efficient daily adaptive proton therapy workflows are being developed [9]. Therefore, the number of adaptations – the adaptation rate – must be limited to what is strictly necessary and should ideally be an optimal trade-off between the clinical gain of each adaptation and the increase in costs associated with additional CT scans and replanning. For a plan to be clinically viable it must be robust against the previously mentioned sources of uncertainty.

Publications like Li et al. linked an increase in robustness with a reduced need for adaptation and thus a lower upfront cost in the form of additional imaging and replanning [10]. However, they also highlighted that regardless of how robust a given plan is, a subset of patients will incur enough anatomical changes during treatment that adaptation will be required. Given the limited amount of literature for proton therapy of cancer patients with tumors located in areas largely impacted by movement such as the thorax, a more in-depth analysis on how a plan's robustness impacts the rate of adaptation could prove beneficial in balancing the limited machine times and patient outcomes [11–17].

The adaptation rate for lung cancer patients undergoing proton therapy has been previously investigated in literature. Chang et al. investigated the rate at which treatment plans are adapted for lung cancer patients using intensity-modulated proton therapy (IMPT) for a 34-patient database with 26.5% of patients requiring adaptation and the median time at which plans were adapted being after 10 fractions during treatment [18]. Hoffmann et al. observed that 61% of their 23-patient database required adaptation on average after 10 fractions with PTV-based plans [6]. This disparity in adaptation rate values can be attributed to a lack of standardization in treatment planning between proton centers (robust optimization, robustness evaluation, imaging modality and frequency of

acquisition, physician protocols, trial design and collection of data etc.) and intrinsic differences in regional patient populations, comorbidities and individual tumor development making the unbiased selection of a database difficult. To our knowledge, no study for NSCLC patients has been done linking the treatment plan's robustness and the rate at which adaptation is performed.

In this publication we aim to estimate the trade-off between plan robustness and the adaptation rate for proton therapy of lung cancer patients.

2. Material and Methods

2.1. Patient data

The planning database contained 40 non-small cell lung cancer patients. Patient data consisted of a planning 4DCT image set and up to eight weekly repeated 4DCTs (reCTs) image sets. All 4DCTs contained eight amplitude-binned breathing phases. The 50% expiration phase was seen as the mid-ventilation CT (MidV-CT), and the MidV-CT of each of the repeated 4DCTs were rigidly registered to the MidV-CT of the planning CT images using the grey level-based image registration tool in RayStation (RaySearch Laboratories) with a focus on bone structures.

A prescription dose of 60 Gy over 30 fractions was used with the goal of delivering at least 95% of the prescribed dose (= 57 Gy) to 98% of the target volume. The target volume was a union of the clinical target volumes (CTVs) for the primary tumor, lymph nodes and metastatic tumors, each created as a 5 mm expansion of the gross tumor volume (GTV). The CTVs were delineated on all eight breathing based on deformable image registration. In the plan optimizations, priority was given to target coverage. OAR constraints, beam angles, number of beams, and number of isocenters were determined on a case-by-case basis, based on tumor size, location, and motion amplitude. Only the CTV coverage was robustly optimized, while the OARs were non-robustly optimized to reduce computation time. The OAR constraints and GTV volumes for each patient can be found in the Appendix (Table S1 and Table S2).

2.2. Robust optimization

Robust optimization was performed in the treatment planning system (TPS) RayStation 12A, on a computer with the following specifications: Intel Xeon Gold 6234 CPU (two 3.30 GHz processors), 128 GB RAM, NVIDIA Quadro RTX 6000 GPU, 2 TB SSD.

All treatment plans used the MidV-CT as the nominal planning CT with two additional breathing phases used for robust optimization: the maximum exhale and inhale CTs (defined as 0% and 100% inhalation phase). A range error value of 2.6%, stemming from the conversion of the CT numbers to SPRs, was used as per reviewing Paganetti et al. [3]. Robustness levels were defined based on an isotropic setup error value of 3, 6 and 9 mm.

When using a setup error margin greater than 5 mm in RayStation 12A, a significant increase in optimization time is observed due to the system automatically generating intermediate error

scenarios to ensure robust coverage. To overcome this limitation, instead of optimizing on the CTV in the 6- and 9-mm setup error plans, a 1- and 4-mm CTV expansion was generated and used in conjunction with a 5 mm isotropic setup error, the sum of which equal our initial margin values [19]. Optimization was thus performed directly on the CTV for the 3 mm plans, and on the expanded CTVs (CTV_1mm and CTV_4mm) for the 6- and 9-mm plans, respectively.

The total number of scenarios used for robust optimization was 63: 7 (setup errors in x,y,z directions and the nominal scenario) × 3 (range errors: ±2.6%, 0%) × 3 (breathing phases).

2.3. Robustness evaluation

Robustness evaluation was performed by generating error scenarios from combinations of maximum errors of each uncertainty source such as setup, range, and breathing motion. Unlike in the case of robust optimization, all eight breathing phases in each given 4DCT image set, planning or repeated, were used for robustness evaluation. A range error value of 2.6% was used. For the treatment plans on the planning CT, robustness evaluation was done on the CTV using the total setup uncertainty values of 3, 6 and 9 mm for the planning CTs (pCTs). For every repeated CT, the original treatment plan was evaluated with the same strategy, but using an isotropic 2 mm setup error, like the approaches of Ribeiro et al. and Visser et al. [20,21].

The total number of scenarios used for robustness evaluation was 168: 7 (setup errors in x,y,z directions and the nominal scenario) × 3 (range errors: ±2.6%, 0%) × 8 (breathing phases).

2.4. Adaptation rate

The adaptation rates of the 3-, 6- and 9-mm robustness levels were defined as the percentage of patients which failed to meet our clinical objectives either in terms of target coverage or OAR overdosing at any point during treatment, for any given reCT set.

As previously mentioned, target coverage was prioritized alongside sparing of critical serial organs (spinal cord). The esophagus structure often overlapped with the CTV which made delivering the prescription dose to a sizeable percentage of its volume unavoidable. While sparing of parallel/mixed organs (lungs-GTV and heart) was considered as much as possible, the particularities of each individual case in terms of larger tumor sizes and larger motion amplitudes also inevitably led to a higher D_{mean} and $V_{30\%}$, the latter often exceeding the clinical threshold of 20%. Given these factors, adaptation rate was determined on whether the following dose metrics were met: CTV $D_{98\%} \geq 57$ Gy and spinal cord $D_{2\%} < 40$ Gy, both based on the worst-case scenario.

2.5. Normal tissue complication probability (NTCP)

Calculation of the normal tissue complication probability (NTCP) values was done based on the Dutch indication protocol for proton therapy, for esophageal and pulmonary toxicity grade 2 or higher (esophagitis and pneumonitis) and mortality 2 years post-treatment [22]. The following formula was used to determine each NTCP value:

$$NTCP = \frac{1}{1+e^{-S}} \quad (1)$$

For esophageal toxicity grade 2 or higher, the value of S was calculated as such:

$$S = 3.634 + 1.496 * \ln(MED) - 0.0297 * (overall\ treatment\ time) \quad (2)$$

Where (MED) is the mean esophageal dose (Gy) and the overall treatment time is the time interval measured in days from the start to the end of the treatment. Here, a fixed 40 day overall treatment time was assumed for every patient.

For pulmonary toxicity grade 2 or higher, the value of S was calculated as such:

$$S = -4.12 + 0.138 * MLD - 0.3711 * (smoking: stopped) - 0.478 * (smoking: active) + 0.8198 * (pulmonary\ comorbidities) + 0.6259 * (tumor\ location) + 0.5068 * (patient\ age) + 0.47 * (sequential\ chemotherapy) \quad (3)$$

Where MLD is the mean lung dose (Gy), (smoking: stopped) is 1 if the patient quit smoking and 0 if the patient is an active smoker or is not a smoker, (smoking: active) is 1 if the patient is an active smoker and 0 if the patient quit or has never smoked, (pulmonary comorbidities) is 1 if the patient has a pre-existing lung disease and 0 if not, (tumor location) is 1 if the tumor is located in the middle or lower lobes and 0 if it is located in the upper lobe, (age) is 1 if patient is 63 years old or older and 0 if younger, and (sequential chemotherapy) is 1 if the patient is undergoing it and 0 if not. Here, concurrent chemotherapy was assumed for every patient.

For mortality 2 years post-treatment, the value of S was calculated as such:

$$S = -1.3409 + 0.0590 * \sqrt{(\mathrm{GTV})} + 0.2635 * \sqrt{(\mathrm{MHD})} \quad (4)$$

Where (GTV) is the tumor GTV volume (cm^3) and (MHD) is the mean heart dose (Gy).

3. Results

The results obtained for the adaptation rate of each robustness level based on the reason for failure (either target $D_{98\%}$ < 57 Gy in the worst-case scenario or spinal cord overdosing $D_{2\%}$ ≥ 40 Gy in the worst-case scenario) can be found below in Table 1 alongside a view on the percentage of failures that occurred as soon as the first reCT. A more in-depth view of the target dose at the point of failure based on $D_{98\%}$ worst-case can be found below in the Supplementary Material (Figure S1) showing that for the 3,6 and 9 mm robustness levels, 29.4%, 20.0% and 31.6% respectively failed by 1 Gy or less. Detailed dose metrics for each patient's first point of failure can be found in the Supplement (Table S2) in the form of either the target $D_{98\%}$ worst-case value or spinal cord $D_{2\%}$ worst-case value as well as the GTV volumes for all patients. The mean gains in OAR sparing, meaning the difference in OAR dose between two given robustness levels, or by what amount the dose received by the OAR is lowered at lower robustness levels can be seen in Table 2 and in the Supplement (Figures S2-S5). The results of the NTCP calculation, the likelihood of incurring pneumonitis and esophagitis grade ≥2 and mortality 2 years post-treatment based on obtained dose metrics and patient data, can be seen in Table 3 and better visualized in Figures 1-3. The NTCP values for each patient can be found in the Supplement (Table S3).

OAR dose metrics based on the pCT were used alongside clinical factors to calculate the NTCP for esophagitis, pneumonitis grade 2 or higher, and mortality 2 years post-treatment. Based on the difference in NTCP between 9 mm and 3 mm cases, an NTCP gain of 2.9% for pneumonitis grade ≥2, 7.6% for esophagitis grade ≥2 and 2.5% for mortality 2 years post-treatment was determined. This should be considered within the planning workflow to improve patient toxicity risks post-treatment when target coverage can be maintained.

The NTCP gains from the 9 mm robustness level to the 3 mm robustness level can be further divided into four segments each representing 25% of patients, based on the gain in each NTCP metric, with the top 25% having the most gain in NTCP between robustness levels and the bottom 25% the lowest NTCP gain from the robustness reduction. The goal of this division is highlighting individual subgroups in our patient database due to the variables in planning lung cancer cases, such as tumor size, location and motion making a reduction in robustness, when possible while maintaining target coverage, significant in some cases but negligible in others. Figures S6-S8 in the Supplement highlight these gains for pulmonary toxicity, esophageal toxicity, and mortality 2 years post-treatment, respectively.

Table 1. Adaptation rate of the 40 patients for each robustness level at any time point with the reason of failure, and the percentage of the patients needing an adaptation that failed on the first repeat CT (reCT).

Setup uncertainty	Adaptation rate (%)	# of patients that required adaptation due to target under-dosage	# of patients that required adaptation due to spinal cord overdose	% of failures that occurred on the first reCT
3 mm	87.5	34	1	74.3
6 mm	70.0	25	3	42.9
9 mm	57.5	19	4	30.4

Table 2. Mean gains (± standard deviation) in D_{mean} (Gy), $D_{2\%}$ worst-case (Gy) and V_{30} (%) on the planning CT between levels of robustness.

Dose metric	9→6 mm	6→3 mm	9→3 mm
Lungs-GTV D_{mean} (Gy)	1.0 ± 0.5	1.1 ± 0.5	2.1 ± 0.9
Lungs-GTV V_{30} (%)	1.5 ± 1.0	1.7 ± 0.9	3.2 ± 1.7
Heart D_{mean} (Gy)	0.8 ± 0.6	0.8 ± 0.6	1.6 ± 1.2

Heart V_{30} (%)	1.3 ± 1.2	1.3 ± 1.1	2.6 ± 2.2
Esophagus $D_{2\ worst\text{-}case}$ (Gy)	2.7 ± 3.8	3.3 ± 3.7	6.0 ± 6.9
Esophagus D_{mean} (Gy)	3.5 ± 0.9	4.1 ± 1.3	7.6 ± 2.2
Spinal cord $D_{2\%}$ (Gy)	4.5 ± 4.2	4.2 ± 3.4	8.7 ± 7.2

Table 3. Mean gains (± standard deviation) in NTCP (pp) on the planning CT between levels of robustness.

NTCP (%)	Δ9-6mm (pp)	Δ6-3mm (pp)	Δ9-3mm (pp)
pulmonary toxicity (pneumonitis grade ≥2)	1.4 ± 1.1	1.5 ± 1.3	2.9 ± 2.2
esophageal toxicity (esophagitis grade ≥2)	3.5 ± 2.0	4.1 ± 2.8	7.6 ± 4.5
mortality risk 2 years post-treatment	1.2 ± 0.6	1.3 ± 0.7	2.5 ± 1.1

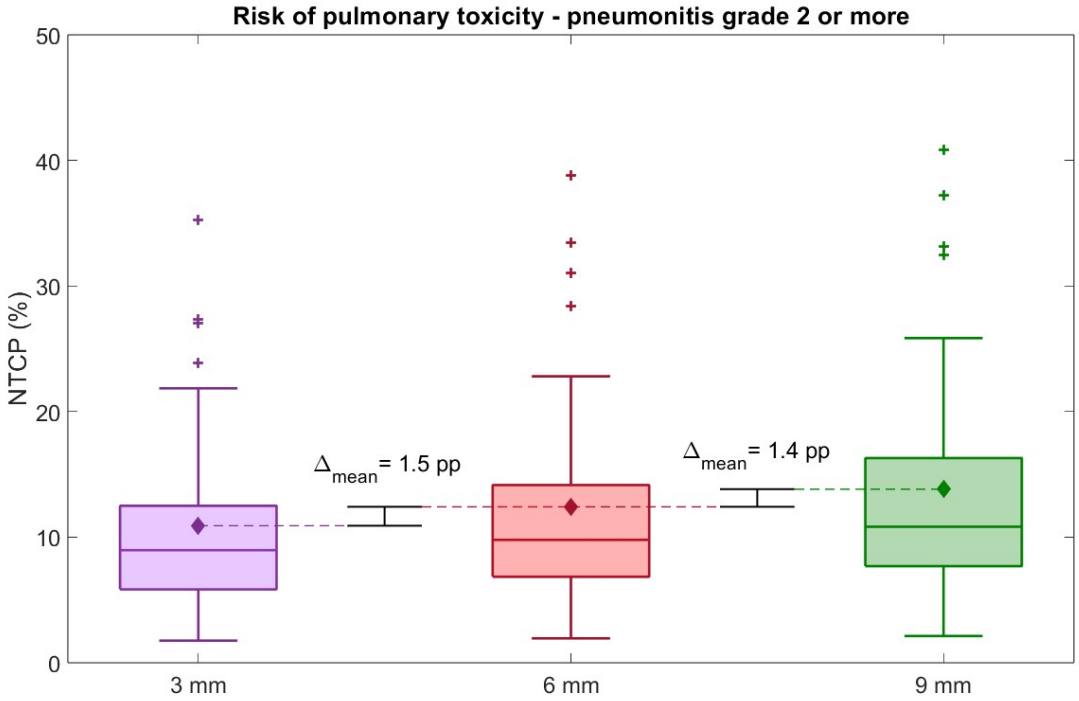

Figure 1. Boxplots and outliers (+) for the NTCP (%) values for pneumonitis grade 2 or higher. The diamonds indicate the mean over the 40 patients, and the Δ_{mean} indicate the differences in the means, signifying the gains in risk of pulmonary toxicity from higher to lower robustness levels.

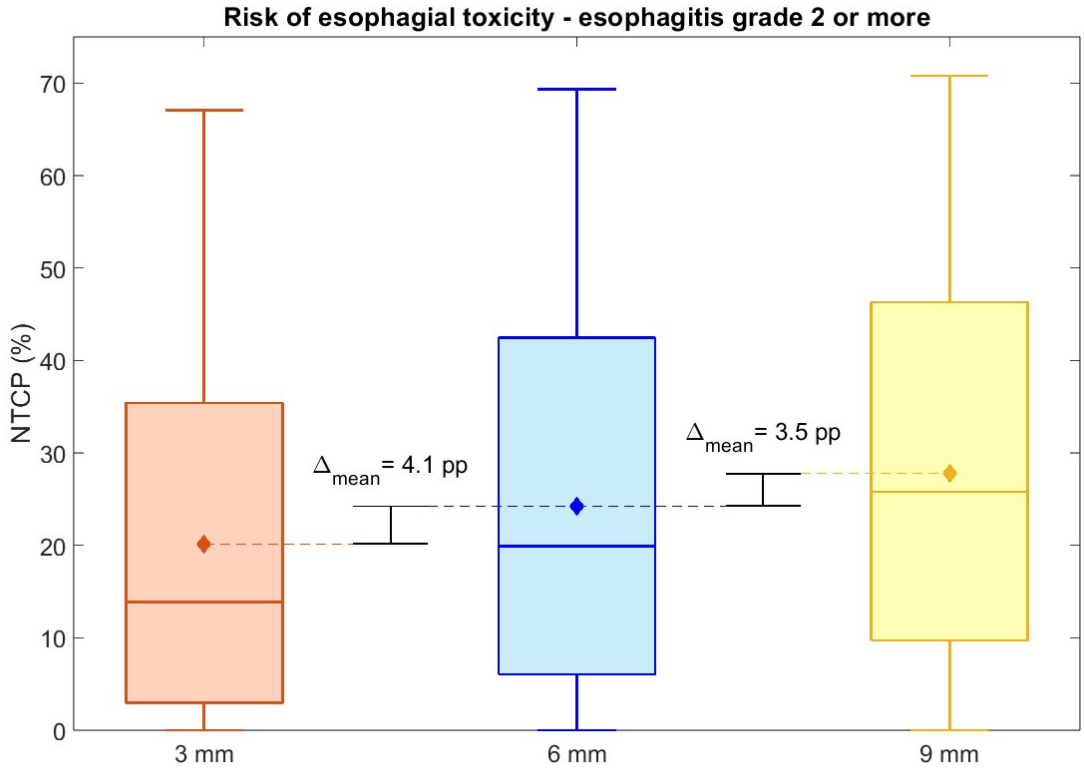

Figure 2. Boxplots for the NTCP (%) values for esophagitis grade 2 or higher. The diamonds indicate the mean over the 40 patients, and the Δ_{mean} indicate the differences in the means, signifying the gains in risk of esophageal toxicity from higher to lower robustness levels.

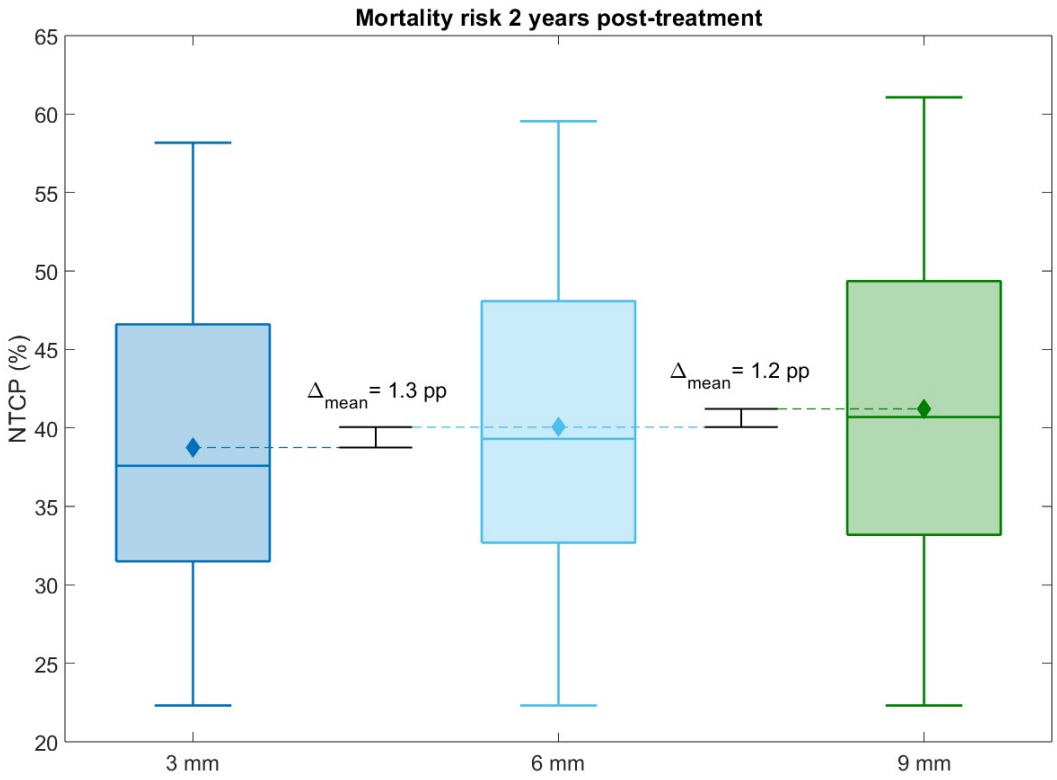

Figure 3. Boxplots for the NTCP (%) values for mortality risk 2 years post-treatment. The diamonds indicate the mean over the 40 patients, and the Δ_{mean} indicate the differences in the means, signifying the gains in mortality risk 2 years post-treatment from higher to lower robustness levels.

4. Discussion

We aim to assess the impact that a treatment plan's robustness has on the adaptation rate for NSCLC patients and what this means for their post-treatment toxicity risk.

Gains in terms of OAR sparing were recorded for all metrics (D_{mean}, D_2, V_{30}) when moving from higher robustness levels to lower ones. This was to be expected since a dose that is more conformal to the target is correlated with improve sparing of OARs at the expense of an increased susceptibility to inter-fractional anatomical changes. The latter issue is evidenced by the increase in adaptation rate at lower robustness levels (Table 1). Out of the 40 patients, only 4 (10%) passed our clinical thresholds for all robustness levels and all reCTs, while 5 patients (13%) only passed on all reCTs for 9 mm plans, and 8 patients (20%) passed on all reCTs only for 9 and 6 mm plans. 22 patients (55%) did not pass for any robustness level, and one (2.5%) only passed in the 3 mm plan (this patient failed on the 6- and 9-mm plan due to spinal cord overdose). The two factors that led to requiring adaptation were target under-dosage ($D_{98\%}$ worst-case < 57 Gy) and spinal cord over-dosage ($D_{2\%}$ worst-case ≥ 40 Gy). No treatment plan had a mean dose for the heart and lungs-GTV that exceeded our clinical threshold of 20 Gy in the nominal scenario. Due to the patient database having a variety of tumors in terms of size, location,

motion amplitude and number and location of lymph nodes and metastatic tumors, keeping V_{30} <20% for heart and lungs-GTV was not always feasible, particularly at the higher robustness levels. Additionally, the particularities of each individual patient led to the esophagus often overlapping with the expanded CTV or, particularly for the 6 and 9 mm cases, the spinal cord being overdosed in the worst-case $D_{2\%}$. The overlap of the esophageal structure with larger volumes receiving the prescription dose resulted in the optimizer trying to ensure that the dose constraint ($D_2 \leq 60$ Gy) is met at higher robustness levels, and in some cases, to deliver a lower $D_{2\%}$ worst-case dose than at lower robustness levels.

It is important to note that even though the adaptation rate numbers we obtained: 87.5%, 70.0% and 57.5% for the 3, 6 and 9 mm robustness levels, respectively, are higher that what is commonly found in literature [6,18], this can be attributed to differences in the methodology of planning and the clinical thresholds used for evaluating. More complex plans involving multiple beam isocenters were assessed by clinicians at the planning CT (pCT) stage, however due to the size of the database and the number of planning doses recomputed on reCTs, the plans reevaluated on the reCTs were not individually assessed for clinical validity by a medical professional and instead a hard dose threshold was established for the target (at least 98% of prescription dose to 95% of tumor volume in the worst-case scenario) and spinal cord (< 40 Gy in the $D_{2\%}$ worst-case scenario) with everything that fails to meet said thresholds being considered unviable. In a realistic clinical setting, several of the plans noted here as failures to meet our clinical goals, particularly by 1 Gy or less for target coverage, representing 29.4%, 20.0% and 31.6%, respectively, for 3-, 6- and 9-mm cases, as seen in Figure S1, could be acceptable when recomputed on the reCTs. As such, these results may represent a more conservative view on adaptation rate and how robustness can impact it instead of a direct parallel to a clinical environment. Nonetheless, the post-treatment benefit to the patient at lower robustness levels in the form of OAR sparing, particularly for cases where target coverage can be maintained remains clinically relevant.

Besides the clinical indicators such as smoking, additional pulmonary comorbidities, patient age, tumor location and sequential chemotherapy, the mean lung dose is the main driving factor for the incidence of post-treatment pneumonitis. For the top 25% of patients, this is reflected in sparing of between 3.6-10.2 pp. This patient group is representative of larger tumor sizes, motion amplitudes and widely distributed lymph nodes and metastatic tumors also irradiated with the prescription dose which led to an undesirable overdosing of the lung structures. The opposite is true for the bottom 25% subgroup, representing small tumor volumes, reduced tumor motion and lack of lymph nodes and metastatic tumors.

The subgroup having the 25% highest NTCP gain for esophageal toxicity showcased gains of between 10.5- 19.6 pp. As previously mentioned, this characterizes primarily patients where the esophagus overlaps with the expanded CTV or the dose distribution of the treatment plan especially at higher robustness levels. The bottom 25% subgroup representing patients where the tumor volume was did not overlap with the plan's dose distribution from the esophagus structure.

The top 25% performers for mortality 2 years post-treatment have gains of between 3.4 pp and 4.0 pp with the driving factors for this metric being the GTV volume and mean heart dose. On a case-by-case basis, the clinical importance of the NTCP gains becomes even more evident. This further emphasizes the need for a robust selection process for lung cancer patients undergoing proton therapy. Individual factors such as proximity to critical OARs which translates into the mean dose,

motion amplitude and tumor size correlated with an adequate level of robustness can result in significant gains in the patient's post-treatment quality of life.

5. Conclusions

By evaluating the impact of a treatment plan's robustness on adaptation rate for lung cancer patients and what it means in terms of a patient's post-treatment toxicity risk in the form of a reduction of toxicities, a cost-effective trade off can be identified. Labor-intensive and financially costly adaptations performed by qualified medical personnel via state-of-the-art imaging and replanning tools should be used no more than necessary and limited beam time for proton therapy should be optimized to accommodate as many patients as feasible. This can be achieved by proper patient selection based on clear and identifiable criteria that prioritize patient outcomes, such as NTCP calculations within a comprehensive regional framework.

Lower robustness levels yielded clinically significant sparing for pneumonitis and esophagitis grade 2 or more and mortality 2 years post-treatment. This is particularly relevant for patients where maintaining target coverage at lower robustness is possible. Regardless how robust a treatment plan is to uncertainties, a subset of patients, particularly with tumors in anatomical regions with significant motion such as the thorax, will still require adaptation. As such, plan robustness should be adjusted on a case-by-case basis, prioritizing target coverage and patient outcome, ensuring surrounding OARs receive as low dose as possible thus minimizing the incidence of toxicities.

Funding

Vlad Mihai Badiu is supported by Stichting tegen Kanker (Grant reference number FAF-C/2018/1195).

Gilles Defraene is a postdoctoral fellow of the Research Foundation Flanders (FWO, project 1292021 N).

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors did not use any generative AI and AI-assisted technologies.

References

[1] Wilson RR. Radiological Use of Fast Protons 1946;47:487–91. https://doi.org/10.1148/47.5.487.

[2] 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. https://doi.org/10.1016/j.ijrobp.2017.05.014.

[3] Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. Phys Med Biol 2012;57:R99. https://doi.org/10.1088/0031-9155/57/11/R99.

[4] Kraus KM, Heath E, Oelfke U. Dosimetric consequences of tumour motion due to respiration for a scanned proton beam. Phys Med Biol 2011;56:6563–81. https://doi.org/10.1088/0031-9155/56/20/003.

[5] 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. https://doi.org/10.1016/j.ijrobp.2013.04.009.

[6] Hoffmann L, Alber M, Jensen MF, Holt MI, Møller DS. Adaptation is mandatory for intensity modulated proton therapy of advanced lung cancer to ensure target coverage. Radiotherapy and Oncology 2017;122:400–5. https://doi.org/10.1016/j.radonc.2016.12.018.

[7] Koay EJ, Lege D, Mohan R, Komaki R, Cox JD, Chang JY. Adaptive/nonadaptive proton radiation planning and outcomes in a phase II trial for locally advanced non-small cell lung cancer. Int J Radiat Oncol Biol Phys 2012;84:1093–100. https://doi.org/10.1016/j.ijrobp.2012.02.041.

[8] Szeto YZ, Witte MG, van Kranen SR, Sonke JJ, Belderbos J, van Herk M. Effects of anatomical changes on pencil beam scanning proton plans in locally advanced NSCLC patients. Radiotherapy and Oncology 2016;120:286–92. https://doi.org/10.1016/j.radonc.2016.04.002.

[9] Nenoff L, Matter M, Charmillot M, Krier S, Uher K, Weber DC, et al. Experimental validation of daily adaptive proton therapy. Phys Med Biol 2021;66. https://doi.org/10.1088/1361-6560/AC2B84.

[10] Li H, Zhang X, Park P, Liu W, Chang J, Liao Z, et al. Robust optimization in intensity-modulated proton therapy to account for anatomy changes in lung cancer patients. Radiotherapy and Oncology 2015;114:367–72. https://doi.org/10.1016/j.radonc.2015.01.017.

[11] Nenoff L, Matter M, Jarhall AG, Winterhalter C, Gorgisyan J, Josipovic M, et al. Daily Adaptive Proton Therapy: Is it Appropriate to Use Analytical Dose Calculations for Plan Adaption? Int J Radiat Oncol Biol Phys 2020;107:747–55. https://doi.org/10.1016/j.ijrobp.2020.03.036.

[12] Jensen MF, Hoffmann L, Petersen JBB, Møller DS, Alber M. Energy layer optimization strategies for intensity-modulated proton therapy of lung cancer patients. Med Phys 2018;45:4355–63. https://doi.org/10.1002/MP.13139.

[13] Gorgisyan J, Munck af Rosenschold P, Perrin R, Persson GF, Josipovic M, Belosi MF, et al. Feasibility of Pencil Beam Scanned Intensity Modulated Proton Therapy in Breath-hold for Locally Advanced Non-Small Cell Lung Cancer. Int J Radiat Oncol Biol Phys 2017;99:1121–8. https://doi.org/10.1016/j.ijrobp.2017.08.023.

[14] Borderías Villarroel E, Geets X, Sterpin E. Online adaptive dose restoration in intensity modulated proton therapy of lung cancer to account for inter-fractional density changes. Phys Imaging Radiat Oncol 2020;15:30–7. https://doi.org/10.1016/j.phro.2020.06.004.

[15] Bernatowicz K, Geets X, Barragan A, Janssens G, Souris K, Sterpin E. Feasibility of online IMPT adaptation using fast, automatic and robust dose restoration. Phys Med Biol 2018;63:085018. https://doi.org/10.1088/1361-6560/AABA8C.

[16] Spautz S, Haase L, Tschiche M, Makocki S, Richter C, Troost EGC, et al. Comparison of 3D and 4D robustly optimized proton treatment plans for non-small cell lung cancer patients with tumour motion amplitudes larger than 5 mm. Phys Imaging Radiat Oncol 2023;27:69192. https://doi.org/10.1016/j.phro.2023.100465.

[17] Feng H, Patel SH, Wong WW, Younkin JE, Penoncello GP, Morales DH, et al. GPU-accelerated Monte Carlo-based online adaptive proton therapy: A feasibility study. Med Phys 2022;49:3550–63. https://doi.org/10.1002/MP.15678.

[18] Chang JY, Li H, Zhu XR, Liao Z, Zhao L, Liu A, et al. Clinical implementation of intensity modulated proton therapy for thoracic malignancies. Int J Radiat Oncol Biol Phys 2014;90:809–18. https://doi.org/10.1016/j.ijrobp.2014.07.045.

[19] Badiu V, Souris K, Buti G, Villarroel EB, Lambrecht M, Sterpin E. Improved healthy tissue sparing in proton therapy of lung tumors using statistically sound robust optimization and evaluation. Physica Medica 2022;96:62–9. https://doi.org/10.1016/J.EJMP.2022.02.018.

[20] Ribeiro CO, Visser S, Korevaar EW, Sijtsema NM, Anakotta RM, Dieters M, et al. Towards the clinical implementation of intensity-modulated proton therapy for thoracic indications with moderate motion: Robust optimised plan evaluation by means of patient and machine specific information. Radiotherapy and Oncology 2021;157:210–8. https://doi.org/10.1016/J.RADONC.2021.01.014.

[21] Visser S, O. Ribeiro C, Dieters M, Mul VE, Niezink AGH, van der Schaaf A, et al. Robustness assessment of clinical adaptive proton and photon radiotherapy for oesophageal cancer in the model-based approach. Radiotherapy and Oncology 2022;177:197–204. https://doi.org/10.1016/J.RADONC.2022.11.001.

[22] Landelijk Platform voor Radiotherapie bij Longtumoren en Landelijk Platform voor Protonentherapie. Landelijk Indicatie Protocol Protonen Therapie Longcarcinoom 2019. https://nvro.nl/images/documenten/rapporten/LIPP_longen_final_01122019.pdf.

Supplementary Material

Additional Figures/Tables

Table S1. Target (clinical target volume, CTV) and organ-at-risk (OAR) constraints.

Contour	Objective	Comment
CTV	$D_{98} \geq 57.0$ Gy	95% of prescription dose. In the worst-case scenario.
Spinal Cord	$D_{2\%} < 40$ Gy	In the worst-case scenario
Lungs-GTV	$D_{mean} < 20$ Gy, $V_{30\%} < 20\%$	In the nominal scenario.
Heart	$D_{mean} < 20$ Gy, $V_{30\%} < 20\%$	
Esophagus	$D_2 \leq 60$ Gy	In the worst-case scenario

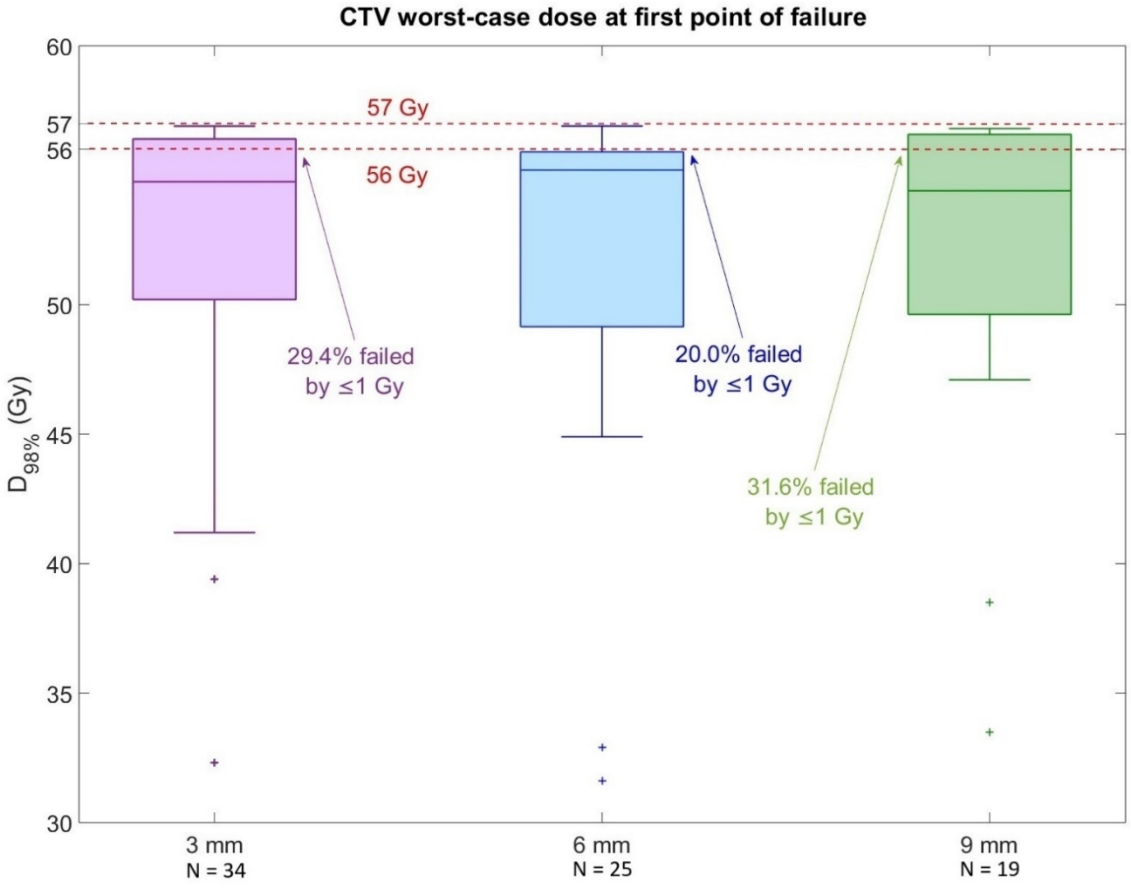

Figure S1. Target $D_{98\%}$ worst-case dose at the first point of failure with 57 Gy dose threshold line and 56 Gy line 1 Gy below our established threshold that may represent a more realistic clinically

acceptable target coverage highlighted in red. Percentage of target under-dosage of 1 Gy or less from minimum clinical threshold shown for each robustness level. N represents the number of patients that failed to meet our clinical threshold for the CTV in the worst-case scenario for each robustness level.

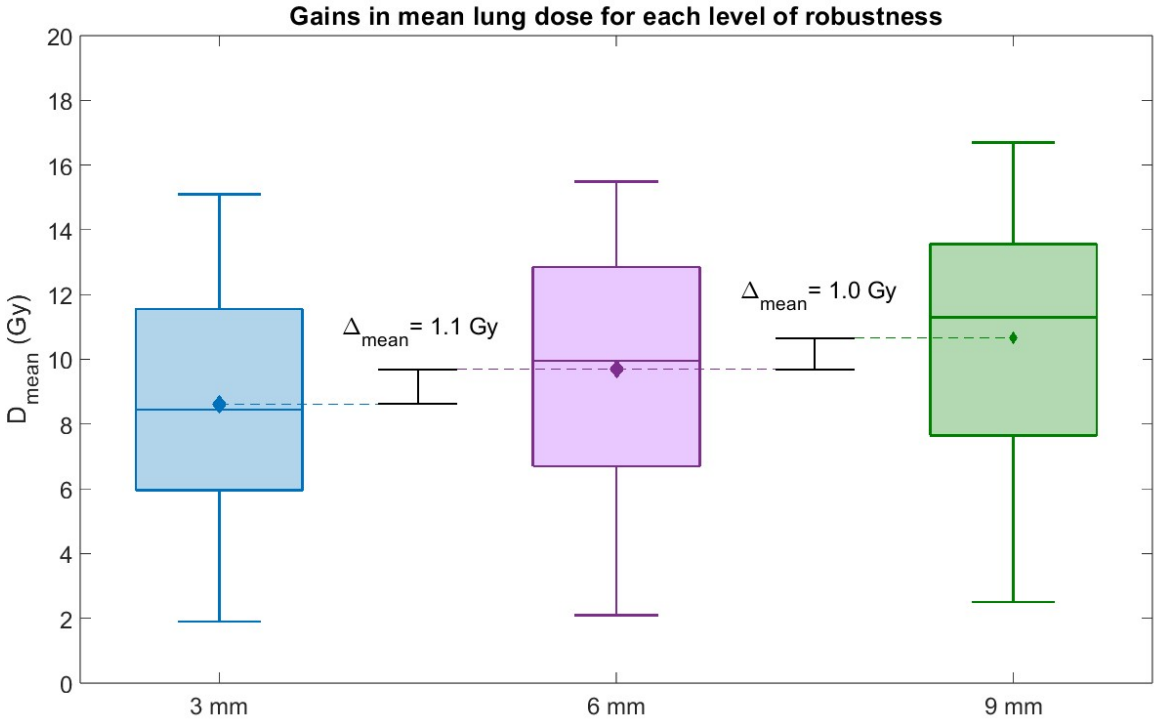

Figure S2. Boxplots for the mean lung dose on the planning CT. The diamonds indicate the mean over the 40 patients, and the Δ_{mean} indicate the differences in the means, or the amount by which the lung D_{mean} (Gy) nominal dose is reduced, on average, from higher to lower robustness levels (gains).

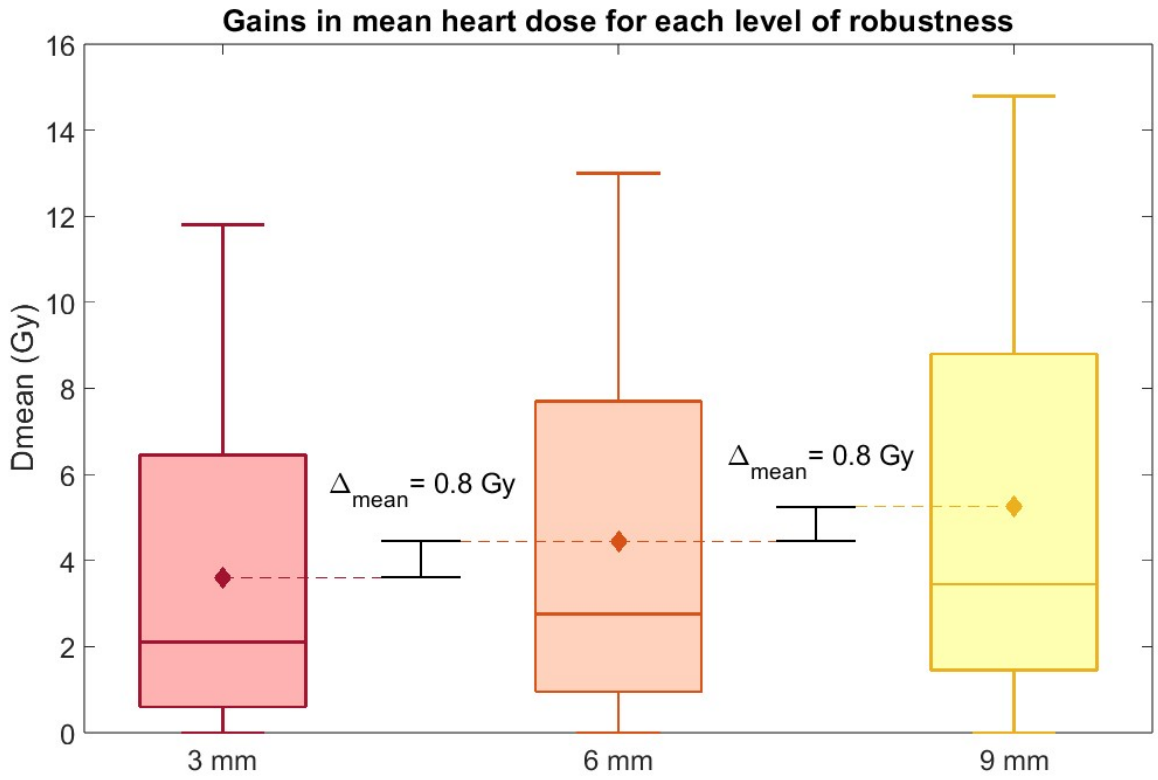

Figure S3. Boxplots for the mean heart dose on the planning CT. The diamonds indicate the mean over the 40 patients, and the Δ_{mean} indicate the differences in the means, or the amount by which the heart D_{mean} (Gy) nominal dose is reduced, on average, from higher to lower robustness levels (gains).

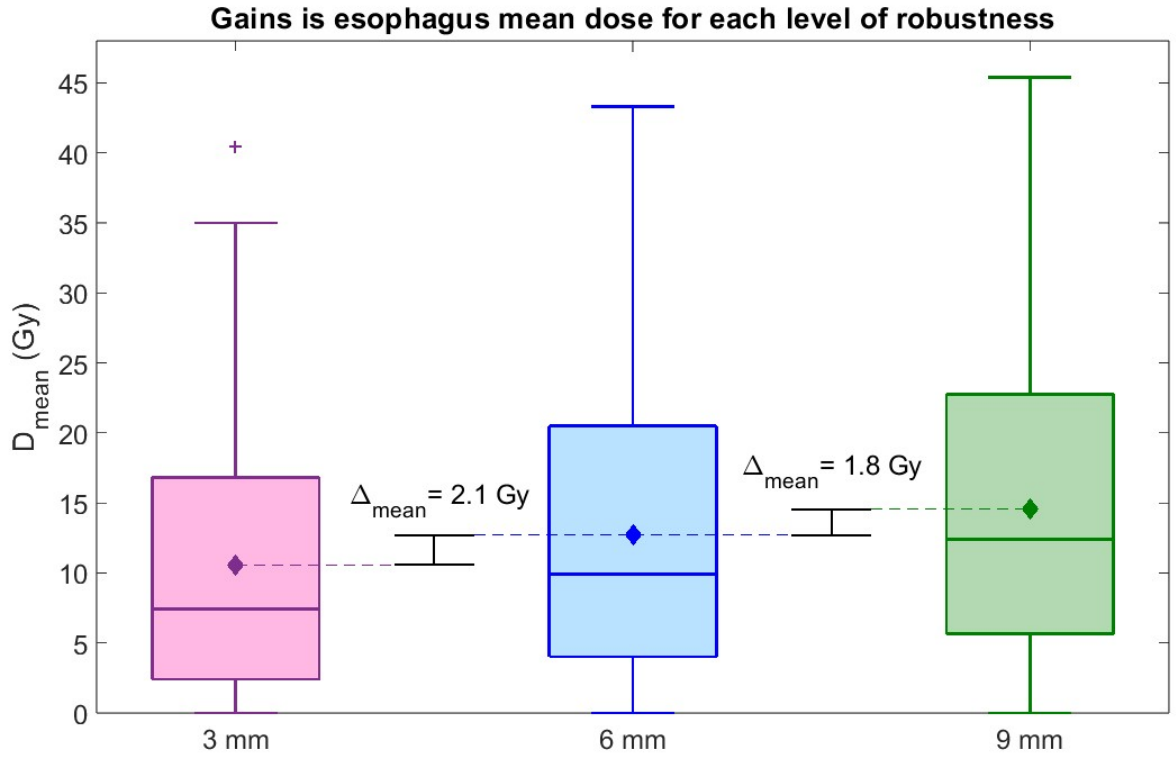

Figure S4. Boxplots and outliers (+) for the mean esophageal dose on the planning CT. The diamonds indicate the mean over the 40 patients, and the Δ_{mean} indicate the differences in the means, or the amount by which the esophagus D_{mean} (Gy) nominal dose is reduced, on average, from higher to lower robustness levels (gains).

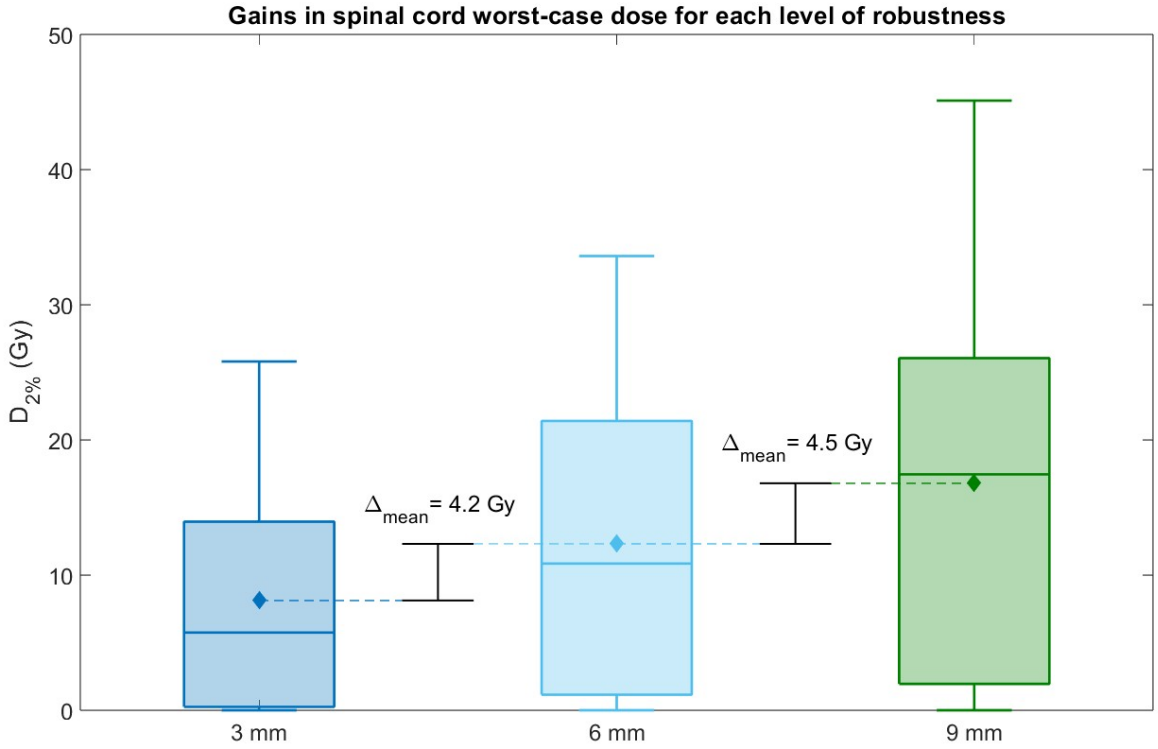

Figure S5. Boxplots for the dose at 2% of spinal cord volume on the planning CT. The diamonds indicate the mean over the 40 patients, and the Δ_{mean} indicate the differences in the means, or the amount by which the spinal cord $D_{2\%}$ (Gy) worst-case dose is reduced, on average, from higher to lower robustness levels (gains).

Table S2. Gross tumor volume (GTV) size for each patient including primary tumor and lymph nodes and dose metrics at the first point of failure. $D_{98\%}$ worst-case scenario failures (<57 Gy) highlighted in red and spinal cord $D_{2\%}$ overdoses (>40 Gy) highlighted in green. Target underdosage is prioritized when reporting a failure, if the target meets our clinical threshold, the spinal cord is then evaluated.

Patient#	GTV volume cm^3	3mm	6mm	9mm
1	331.6	56.4	56.8	
2	326.5	49.9	54.2	44.8
3	307.7		41.4	47.2
4	215.0	56.9		
5	187.3	55.5		
6	173.5	50.2	54.0	56.7
7	168.4	53.1	32.9	33.5
8	109.4	56.5		
9	102.7			
10	101.1	39.4	44.9	51.0
11	89.2	41.2	47.8	49.3
12	79.6			

13	70.0	56.6	56.2	56.6
14	60.8	56.8	56.5	45.1
15	55.4	56.2		
16	55.1	55.2	42.2	45.1
17	54.0	52.6	55.8	56.7
18	52.4	55.7	53.8	54.7
19	49.8			
20	42.3	55.2	55.9	
21	41.4	52.5	55.8	54.4
22	36.4	53.7	56.6	
23	30.0	53.1	55.8	55.7
24	28.0	47.4	54.2	55.1
25	25.8			
26	24.2	53.3	56.9	
27	22.5	55.1	55.2	56.5
28	20.8	49.9	55.1	47.1
29	17.9	42.7	47.5	49.7
30	15.8	55.8		
31	15.0	54.4	55.9	56.7
32	14.7	42.4	49.6	52.9
33	14.4	56.7	55.3	
34	13.2	54.4	50.5	52.0
35	12.4	56.7		
36	12.2	32.3	31.6	38.5
37	9.6	56.0		
38	8.5	56.5	47.2	49.6
39	3.0	50.0	55.3	56.8
40	2.5	56.6		

Table S3. Normal tissue complication probability (NTCP; in %) calculation results for each robustness level.

Patient#	pneumonitis grade ≥2 (%)			esophagitis grade ≥2 (%)			mortality risk 2 years post-treatment (%)		
	3mm	6mm	9mm	3mm	6mm	9mm	3mm	6mm	9mm
1	9.0	9.7	10.9	44.6	49.2	52.3	48.8	50.2	51.9
2	5.4	6.2	7.3	15.3	23.2	29.2	51.5	53.9	55.1
3	6.5	6.9	7.8	20.4	29.2	33.5	47.5	48.9	50.4
4	8.9	9.8	10.1	4.6	6.7	9.3	41.1	41.8	43.2
5	6.0	7.0	7.5	34.4	42.8	48.0	53.5	54.5	55.4
6	16.9	18.3	20.0	67.1	69.3	70.8	58.2	59.5	61.1
7	23.9	28.4	32.5	43.0	48.6	53.0	48.6	50.6	51.9
8	17.1	20.7	25.8	55.1	57.9	60.4	54.5	55.3	56.4
9	11.6	12.0	12.3	17.0	27.9	36.6	52.6	54.4	56.2
10	35.3	38.8	40.8	43.0	46.2	48.3	45.1	46.6	48.2

11	11.1	12.0	12.9	11.9	14.6	18.2	42.1	43.1	43.1
12	27.3	31.0	33.1	55.1	60.6	63.7	48.3	50.3	52.2
13	14.3	15.3	16.2	37.2	42.1	45.7	47.5	49.9	51.4
14	21.8	22.8	23.0	62.2	63.3	64.1	44.7	46.7	48.3
15	10.7	13.2	14.1	36.4	43.4	46.9	45.7	47.2	48.1
16	6.0	6.8	7.7	1.3	2.6	5.8	32.4	32.8	33.2
17	8.1	8.7	9.4	1.2	2.6	5.4	37.4	38.7	40.1
18	11.9	13.9	16.3	0.2	0.7	1.2	45.4	46.5	47.8
19	11.7	14.0	16.3	8.7	13.1	18.2	36.5	38.0	39.2
20	9.1	10.6	12.8	8.7	12.2	16.8	35.4	36.6	37.9
21	5.7	7.1	8.3	31.2	36.6	39.7	38.2	40.2	41.6
22	9.9	12.0	14.0	16.5	24.4	27.6	44.2	46.1	47.6
23	5.2	6.0	6.9	5.4	8.7	14.3	34.2	35.8	37.0
24	7.5	8.3	9.0	1.9	4.6	10.7	30.1	30.8	31.8
25	5.5	7.0	9.3	49.4	53.1	56.5	44.7	46.5	48.1
26	10.8	12.5	14.3	5.0	9.8	12.2	37.8	39.9	41.0
27	4.8	5.5	6.1	0.0	0.0	0.0	25.7	25.7	27.3
28	10.6	12.0	13.3	25.6	31.0	35.8	36.9	38.7	40.4
29	5.1	6.2	7.7	12.4	16.5	24.0	31.9	32.6	33.2
30	3.5	4.3	5.4	20.1	26.0	30.8	28.9	29.8	31.1
31	8.2	9.5	10.8	3.8	6.9	9.8	33.1	33.8	34.5
32	3.2	3.9	4.6	19.9	27.9	34.2	31.6	33.2	34.8
33	8.5	9.6	10.7	8.2	13.4	18.0	29.9	31.4	32.6
34	13.1	14.3	15.3	0.1	0.3	0.3	31.4	33.1	34.7
35	16.0	18.0	19.7	0.6	1.3	2.7	24.4	25.9	25.9
36	4.3	4.4	4.7	2.1	5.4	9.6	24.3	24.3	24.3
37	27.0	33.4	37.2	32.1	40.3	45.1	35.0	36.8	38.7
38	7.1	8.0	8.8	4.2	7.1	11.0	26.4	26.9	27.9
39	1.8	1.9	2.1	0.0	0.0	0.0	22.5	22.5	22.5
40	5.9	6.5	7.2	0.1	0.1	0.4	22.3	22.3	22.3

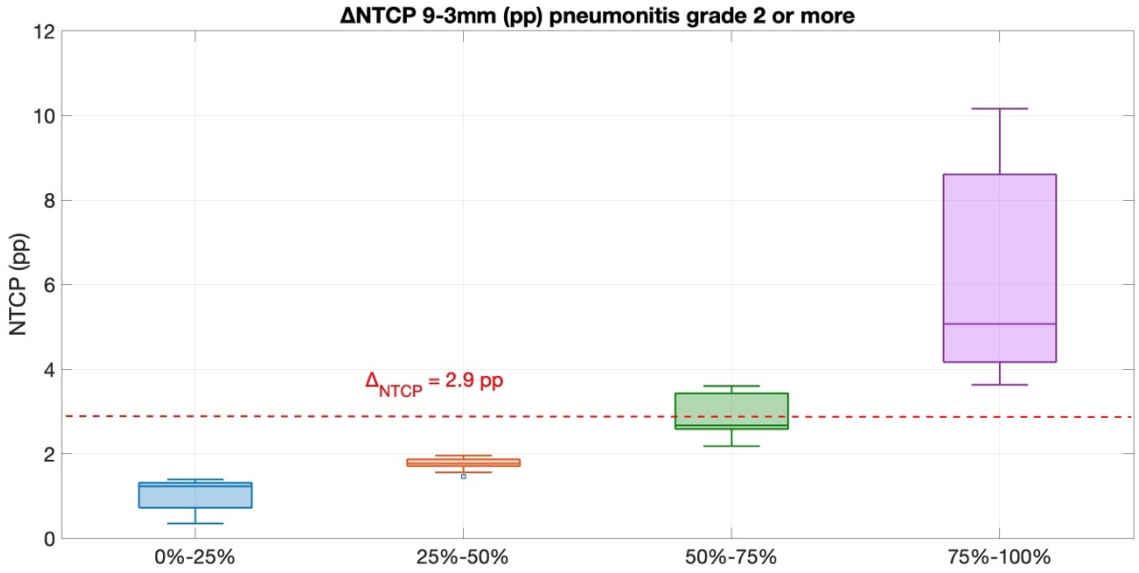

Figure S6. Pulmonary toxicity NTCP gains from 9 mm to 3 mm robustness level divided into four subgroups of 25% of the patients based on the amount of NTCP gain, that is the subgroup denoted 0%-25% is the 25% of the patients having the lowest NTCP gain. The mean gain over all patients is highlighted in red dashed line.

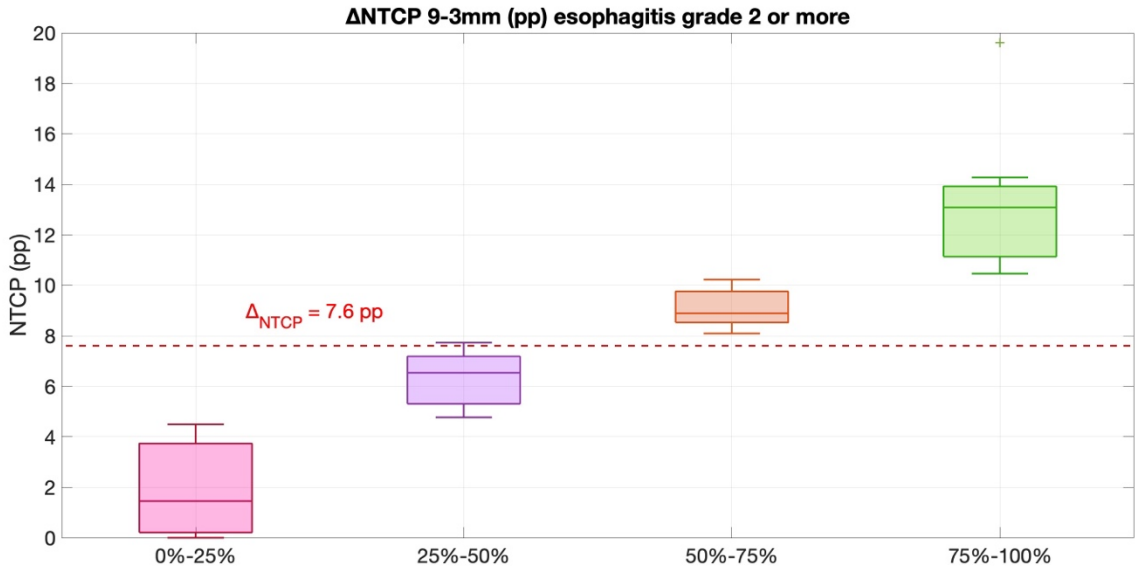

Figure S7. Esophageal toxicity NTCP gains from 9 mm to 3 mm robustness level divided into four subgroups of 25% of the patients based on the amount of NTCP gain, that is the subgroup denoted 0%-25% is the 25% of the patients having the lowest NTCP gain. The mean gain over all patients is highlighted in red dashed line.

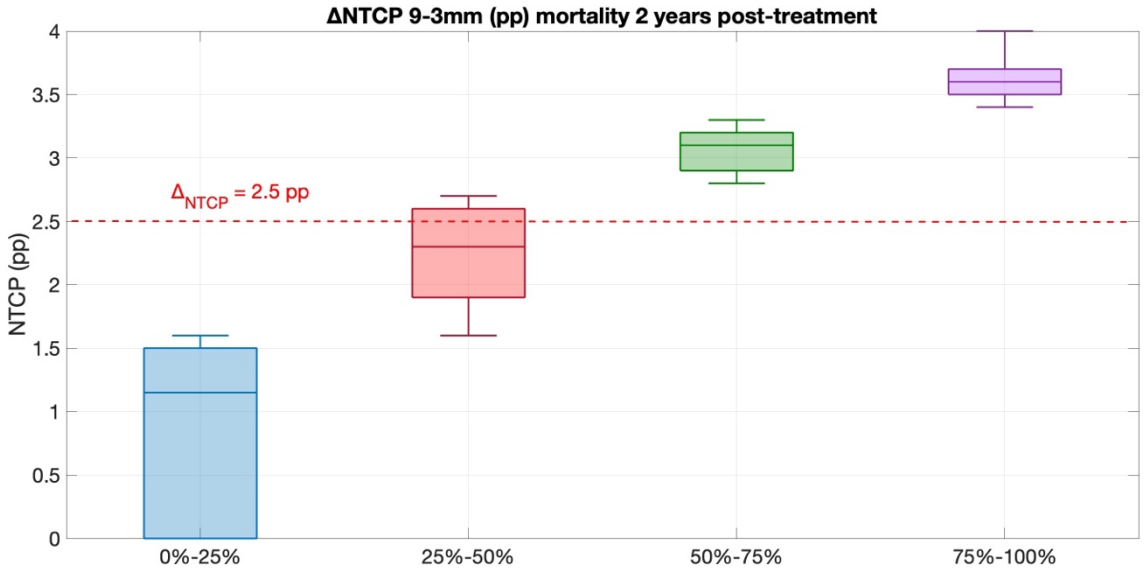

Figure S8. Mortality 2 years post-treatment gains from 9 mm to 3 mm robustness level divided into four subgroups of 25% of the patients based on the amount of NTCP gain, that is the subgroup denoted

0%-25% is the 25% of the patients having the lowest NTCP gain. The mean gain over all patients is highlighted in red dashed line.