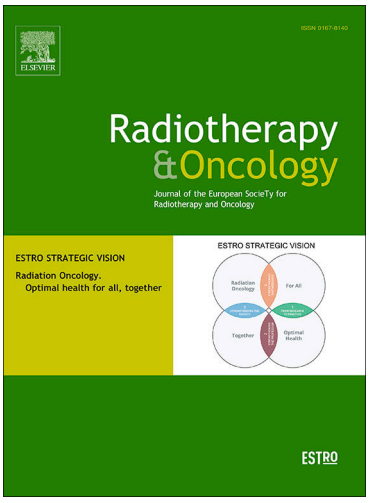
Journal Pre-proofs



Original Article

Inter- and intrafractional 4D dose accumulation for evaluating ΔNTCP robustness in lung cancer

A. Smolders, A.C. Hengeveld, S. Both, R. Wijsman, J.A. Langendijk, D.C. Weber, T. Lomax, F. Albertini, G. Guterres Marmitt

PII: S0167-8140(23)00026-9
DOI: https://doi.org/10.1016/j.radonc.2023.109488
Reference: RADION 109488

To appear in: *Radiotherapy and Oncology*

Received Date: 30 November 2022
Revised Date: 12 January 2023
Accepted Date: 18 January 2023

Please cite this article as: Smolders, A., Hengeveld, A.C., Both, S., Wijsman, R., Langendijk, J.A., Weber, D.C., Lomax, T., Albertini, F., Guterres Marmitt, G., Inter- and intrafractional 4D dose accumulation for evaluating ΔNTCP robustness in lung cancer, *Radiotherapy and Oncology* (2023), doi: https://doi.org/10.1016/j.radonc.2023.109488

Inter- and intrafractional 4D dose accumulation for evaluating ΔNTCP robustness in lung cancer

Smolders A.[a,b], Hengeveld A. C.[c], Both S.[c], Wijsman R.[c], Langendijk J.A.[c], Weber D.C.[a,d,e], Lomax T.[a,b], Albertini F.[a], Guterres Marmitt G.[c]

[a] Centre for Proton Therapy, Paul Scherrer Institute, Villigen, Switzerland
[b] Department of Physics, ETH Zurich, Zurich, Switzerland
[c] Department of Radiation Oncology, University Medical Center of Groningen, University of Groningen, Groningen, The Netherlands
[d] Department of Radiation Oncology, University Hospital Zurich, Zurich, Switzerland
[e] Department of Radiation Oncology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

Abstract

Background and purpose: Model-based selection of proton therapy patients relies on a predefined reduction in normal tissue complication probability (NTCP) with respect to photon therapy. The decision is necessarily made based on the treatment plan, but NTCP can be affected when the *delivered* treatment deviates from the plan due to delivery inaccuracies. Especially for proton therapy of lung cancer, this can be important because of tissue density changes and, with pencil beam scanning, the *interplay effect* between the proton beam and breathing motion.
Materials and Methods: In this work, we verified whether the expected benefit of proton therapy is retained despite delivery inaccuracies by reconstructing the *delivered* treatment using log-file based dose reconstruction and inter- and intrafractional accumulation. Additionally, the importance of two uncertain parameters for treatment reconstruction, namely deformable image registration (DIR) algorithm and α/β ratio, was assessed.
Results: The expected benefit or proton therapy was confirmed in 97% of all studied cases, despite regular differences up to 2 percent point (p.p.) NTCP between the *delivered* and *planned* treatments. The choice of DIR algorithm affected NTCP up to 1.6 p.p., an order of magnitude higher than the effect of α/β ratio.
Conclusion: For the patient population and treatment technique employed, the predicted clinical benefit for patients selected for proton therapy was confirmed for 97.0 % percent of all cases, although the NTCP based proton selection was subject to 2 p.p. variations due to delivery inaccuracies.

Keywords: proton therapy, lung cancer, NTCP, robustness, DIR

Introduction

In the Netherlands, a model-based approach is used to select lung cancer patients for proton therapy [1]. The potential benefit of proton therapy with respect to volumetric modulated arc therapy (VMAT) is quantified by a difference in Normal Tissue Complication Probability (ΔNTCP) [2], determined by a comparison between a clinical photon and proton plan for each patient. If ΔNTCP is larger than a predefined threshold, the patient is selected for proton therapy, else they will be treated with photons. This decision is necessarily made based on the *planned treatment*, but the *delivered treatment* may deviate from the plan because of delivery inaccuracies.

Due to its high precision, proton therapy is particularly sensitive to factors affecting the delivery accuracy [3]. These include errors in beam positioning (machine errors), anatomical changes [4], proton range uncertainty, patient setup errors [5,6] and intrafraction motion (e.g. breathing).

Whereas motion affects the geometrical delivery accuracy independently of the treatment modality, proton therapy is particularly sensitive to it because of tissue density changes. Moreover, most centers employ pencil beam scanning [7], which is additionally affected by the *interplay effect* between the timing of the pencil beam and target motion [8]. For indications strongly affected by breathing motion, such as lung cancer, this may result in hot or cold spots of delivered dose [9].

To verify whether the ΔNTCP based on the planned treatment is robust to delivery inaccuracies, the delivered dose can be reconstructed after the treatment and ΔNTCP can be recalculated. For lung cancer patients, this requires time-resolved dose reconstruction on 4D images based on the machine log files. When repeated for several representative fractions throughout the treatment, these *fraction doses* can be accumulated into the *treatment dose* with a method called 4D Dose Reconstruction and Accumulation (4DREAL) [10].

Previous work based on 4DREAL found no significant loss of target coverage homogeneity due to the breathing motion [11]. However, NTCP is based on dose to organs-at-risk (OARs), including the heart, lungs, esophagus and spinal cord for lung cancer [12]. Whereas several measures are taken to increase target coverage robustness in proton plans [11], this robustness is not designed to extend outside of the target. Therefore, the OAR doses could significantly differ from the planned dose affecting NTCP and this effect has not yet been studied.

Two important factors affect the estimation of NTCP based on accumulated dose. Firstly, accumulating the breathing phase-resolved doses on a common reference CT requires deformable image registration (DIR), a problem that cannot be unambiguously solved with current CT imaging and registration algorithms. A variety of DIR techniques exist in the field, with each potentially performing differently. Therefore, 4DREAL doses depend on the DIR algorithm employed, which impacts the NTCP value of the reconstructed treatment [11,13,14].

Secondly, whereas fractionated treatments are planned to be delivered uniformly in time, daily dose variations are common. Accumulating the physical fraction doses masks these variations: the same accumulated physical dose can be obtained with both uniform and non-uniform fraction doses. Since the biological effect-dose relation is non-linear [15,16], the delivered dose has to be corrected for biological effects before NTCP evaluation. Whereas previous results show that this effect is minimal in high and homogeneous dose regions such as targets [11,17,18], the dose to the OARs is typically low and inhomogeneous, making the daily dose variations relatively larger. Further, the α/β ratio in OARs in lung cancer was found to be low [19–23], increasing the non-linear effect of daily dose variations. Therefore, including non-linear effects in dose accumulation may be important for NTCP evaluation.

This work investigates whether ΔNTCP obtained at the time of planning remains robust to dose delivery inaccuracies and biological effects due to daily treatment variations. ΔNTCP of the delivered dose is retrospectively evaluated for twenty lung cancer patients treated with proton therapy at the University Medical Center Groningen (UMCG) and the anticipated clinical benefit of proton therapy is verified. The effects of DIR uncertainty and potential biological effects of daily dose variations are assessed as both factors affect the delivered NTCP.

Materials and Methods

Patient data

A cohort of 20 lung cancer patients treated with proton therapy at UMCG was retrospectively analyzed. Only patients without plan adaptations were included. Patient anatomy and target motion

were obtained from weekly repeated 4DCT images acquired in treatment position, which were considered representative of the anatomy and breathing pattern for all fractions in that week.

Following clinical standard procedure, treatment plans were robustly optimized on the average planning CT with 6mm setup uncertainty and 3% range uncertainty using Raystation (RaySearch Laboratories AB, Stockholm, Sweden) [24].

4D dose reconstruction and accumulation

The delivered dose was estimated using an adaptation of the log-file based 4D Dose Reconstruction and Accumulation (4DREAL) method [10,25]. Dose reconstruction requires time-resolved patient imaging (4DCT), delivery log files and an ANZAI pressure belt breathing signal (ANZAI Medical, Tokyo, Japan), acquired during treatment delivery and 4DCT acquisition (**Error! Reference source not found.**). The ANZAI breathing signal was used to match the 10 4DCT breathing phases to the correct time during delivery. The collection of spots present in the log files were split by delivery time and converted into *subplans* specific to the 4DCT phases. The Monte Carlo simulation software MCsquare (http://openmcsquare.org/) recalculated each subplan on the corresponding CT.

The time-resolved dose maps corresponding to each 4DCT phase were consequently warped using deformable image registration (DIR) to the 50% inhale phase of that fraction and accumulated into the delivered fraction dose d_f. These fraction doses were then accumulated after warping them by DIR to the average planning CT, on which target and OARs are delineated.

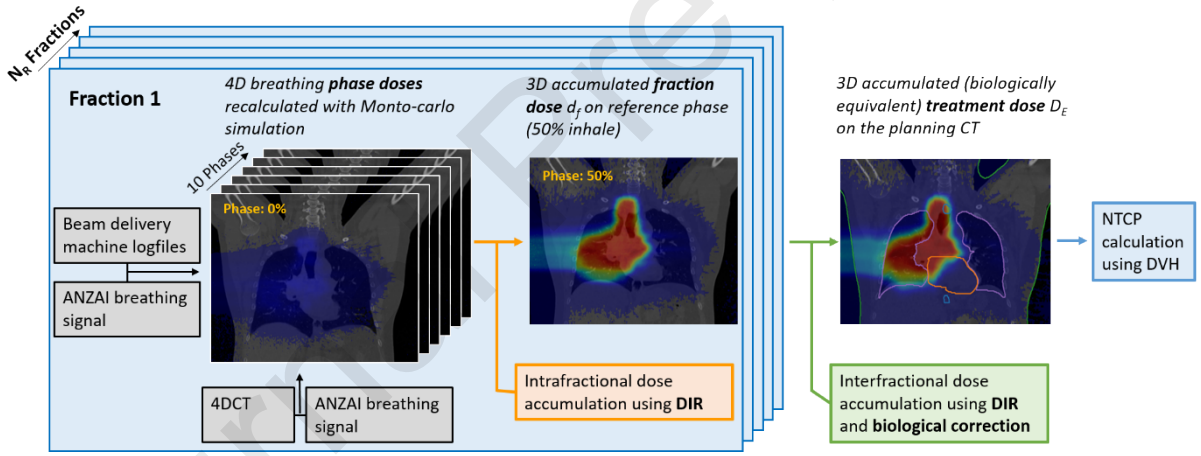


Figure 1: Overview of the 4D Dose Reconstruction and Accumulation (4DREAL) workflow.

Deformable Image Registration uncertainty

Since DIR solutions depend on the employed algorithm, the accumulated dose differs depending on the DIR algorithm. If well-defined anatomical markers were available, the most accurate DIR algorithm could be found for each pair of scans [13,26,27] . However, such data was not available in this project.

Instead of using one DIR algorithm, we used five substantially different ones and accumulated the dose five times for each patient to evaluate the sensitivity of the reconstructed $\Delta\mathrm{NTCP}$. The DIR algorithms and their implementation details can be found in the supplementary material. To focus on the differences between the deformable part of the registration, the initial rigid alignment between the planning CT and reference phase was done with RayStation and applied for all algorithms.

The differences between the DIR algorithms can be random and systematic. When accumulating the dose over multiple fractions, random differences cancel out and systematic ones persist. Systematic

differences can occur, for example, if one DIR is more regularized than another, causing it to systematically yield smaller deformations in regions with large deformation gradients. Here, the accumulation is based on *weekly* 4DCTs, so random differences cancel out less than if 4DCTs were available for each *daily* fraction. This, however, does not affect the systematic difference. To extrapolate our findings to the case where one 4DCT per fraction is available, the relative magnitude of the random and systematic errors is assessed by comparing the DIR differences found when accumulating over one single fraction with the difference found when accumulating over all available 4DCTs.

Biological effect of daily dose variation

Due to anatomical, set-up, range and interplay uncertainties, the delivered dose varies from day to day. When accumulating the doses, the daily dose variations can cancel out, i.e. an irregular dose delivery (in time) can lead to the same accumulated physical dose as a homogeneous one. However, the biological effect would be different, because of the non-linear dose-effect relation.

The biological effect can be accounted for by calculating the *biologically equivalent dose* D_E, i.e. the hypothetical physical dose delivered uniformly over N fractions which has the same biological effect as the non-uniform treatment [28]. Given several representative fractions N_R with corresponding fraction dose d_f, D_E is calculated as:

$$D_E = N\, d_E,$$

$$d_E = -\frac{\alpha}{2\beta} + \sqrt{\frac{\alpha^2}{4\beta^2} + \frac{1}{N_R}\sum_{f=1}^{N_R}\left(\frac{\alpha}{\beta} d_f + d_f^2\right)},$$

with $d_E$the biologically equivalent uniform fraction dose and α and β the parameters of the linear-quadratic model [29]. In this work, the number of delivered fractions N was between 25 and 30 and N_R was the number of repeated 4DCTs, on average $N_R = 4$.

$\mathrm{D_E}$ depends on the parameter α/β, a tissue-specific quantity which determines the relation between dose and biological effect. In target tissues for lung cancer, $\alpha/\beta = 10$ Gy is standardly assumed [30]. For OAR tissue, α/β is expected to be lower, e.g. between 2 and 3 Gy for the heart [19–21] and between 1.3 and 4.5 normal lung tissue [22,23]. In esophagus, $\alpha/\beta = 10$ Gy has been used for acute toxicity [31], but no consistent value is given in the literature.

As the exact value of α/β is unknown and may affect NTCP, we vary α/β with values 0, 2, 3, 10 and ∞. $\alpha/\beta = 0$ depicts a quadratic dose-effect relation, $\alpha/\beta \to \infty$ depicts a linear relation, i.e. the physical dose. Note that this only affects the interfractional dose accumulation, within a fraction the doses are accumulated physically because the delivery is too fast to be impacted by these biological effects.

NTCP impact

As the aim is to evaluate the robustness of ΔNTCP *based on the planned treatment dose* ($\Delta\mathrm{NTCP_{plan}}$), we need to convert D_E into NTCP to estimate ΔNTCP *based on the delivered treatment dose* ($\Delta\mathrm{NTCP_{delivered}}$). D_E further depends on both α/β and DIR, which hence also affect NTCP. Therefore, for each patient, α/β and DIR, we evaluate a voxel-wise D_E on the average planning CT and use the OAR contours to evaluate the corresponding NTCP values. First, we use this to estimate the influence of α/β and DIR on NTCP, after which the values are compared to the planned NTCP.

NTCP predictions were considered robust when $\Delta\mathrm{NTCP_{delivered}}$ did not deviate enough from $\Delta\mathrm{NTCP_{plan}}$ to disqualify the patient from being treated with protons. Note that we only study robustness

for patients originally selected for proton therapy and exclude patients selected for photons, even though they might have benefitted from protons based on accumulated dose. Lung cancer patients qualify for proton treatment when it results in a difference of ≥2 percent point (p.p.) in 2-year-mortality NTCP or ≥10 p.p. for Grade ≥2 acute dysphagia or radiation pneumonitis when compared to the VMAT photon treatment. NTCP calculation is based, among other patient-specific factor on the mean dose to respectively the heart, esophagus and lung minus gross tumor volume (GTV)[12].

Results

The variability of NTCP with respect to the α/β ratio is small (Figure 2). The difference between using the clinical reference $\alpha/\beta = 3$ and any other value remains below 0.16 p.p. for 95% cases, and is usually much lower. For one patient, the difference in radiation pneumonitis NTCP was as large as 0.55 p.p., but only for fully quadratic dose accumulation ($\alpha/\beta = 0$). Without $\alpha/\beta = 0$, the maximal difference is 0.18 p.p. Even though D_E can locally vary significantly, the variations are too concentrated to influence the average dose to the OARs, and hence NTCP, substantially. This means that the daily dose variations do not significantly affect the side effects of the treatment, similar to what was found for the target doses.

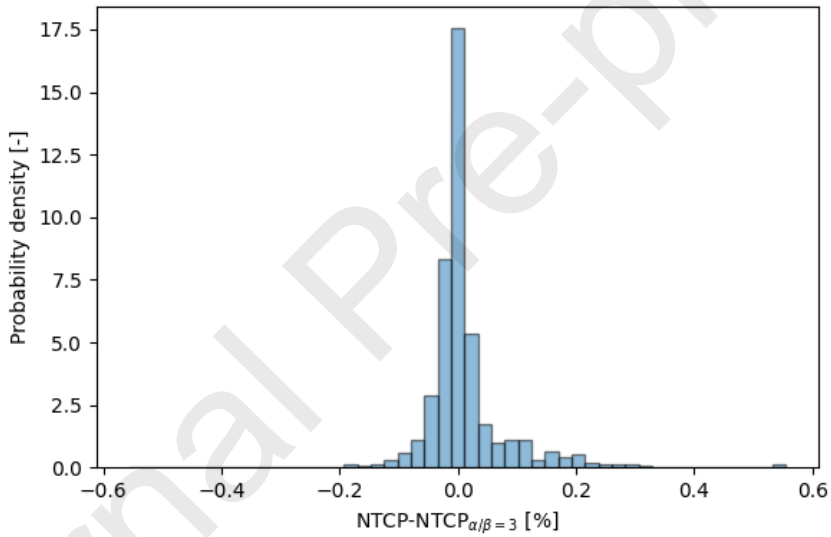


Figure 2: Distribution of the difference in NTCP for several values of α/β ratio with the NTCP for the clinical reference ratio α/β = 3. The differences are calculated for each patient, each DIR algorithm and each endpoint individually and are all included in the histogram.

The influence of the DIR algorithm on the NTCP values is approximately one order of magnitude larger than that of α/β ratio (Figure 3). Accumulating the dose with a different DIR than the clinical reference can change the NTCP up to 1.6 p.p. (except for one patient, see further). This value is close to the 2-year mortality threshold of 2 p.p. ΔNTCP for patient selection for proton therapy, meaning that it cannot be neglected when verifying the robustness of $\Delta\mathrm{NTCP}_{\mathrm{plan}}$.

For one patient, the difference in dysphagia probability between the DIRs is 3.7 p.p. This can be attributed to the combination of a large dose gradient on the border of the esophagus and a difference in the esophagus filling (air) between the planning and repeat CTs causing the DIR solutions to deviate from each other. Even though this is exceptional, it shows that dose accumulation with a single DIR should be carefully interpreted, especially for small organs such as the esophagus.

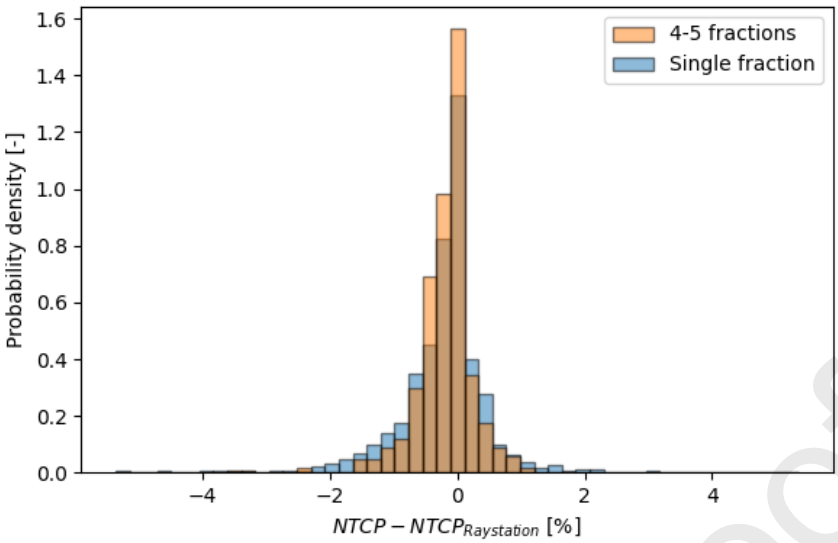


Figure 3: Distribution of the difference in NTCP for each DIR with the NTCP for the clinical reference RayStation DIR. The differences are calculated for each patient, each α/β ratio and each endpoint individually and are all included in the histogram. The orange histogram depicts the situation with 4-5 fractions taken as representative for the daily dose variations. The blue histogram depicts the situation if only a single fraction is taken to represent the treatment.

When accumulating over one representative fraction, the variation between the DIRs is higher than accumulating over 4-5 representative fractions because of the random deviations cancelling out more (Figure 3). If the differences between the DIRs were fully random, the standard deviation σ between the two cases would scale with $\sqrt{1/N} \sim \sqrt{1/4} = 1/2$. However, σ only drops from 0.70 to 0.47, which indicates that part of the differences in DIR are systematic. This further implies that accumulation over all 25 fractions would reduce the DIR differences with less than $\sqrt{N_R/N} \sim \sqrt{4/25} = 2/5$.

The robustness of $\Delta\mathrm{NTCP}_{\mathrm{plan}}$ is evaluated by comparing it to $\Delta\mathrm{NTCP}_{\mathrm{delivered}}$ for all α/β ratios and all DIRs (Figure 4). Note that a fair comparison requires that target coverage objectives are met in both cases, which was verified for all patients (Supplementary material B.), in line with previous results [11]. Figure 4 shows that $\Delta\mathrm{NTCP}_{\mathrm{plan}} > 0$ for all endpoints and all patients. After delivery, this still holds for 97.0% of all cases. Further, we find that all patients selected for protons based on $\Delta\mathrm{NTCP}_{\mathrm{plan}} \geq 2$ p.p. for 2-year mortality or $\Delta\mathrm{NTCP}_{\mathrm{plan}} \geq 10$ p.p. for radiation pneumonitis or dysphagia would still qualify for protons using $\Delta\mathrm{NTCP}_{\mathrm{delivered}}$.

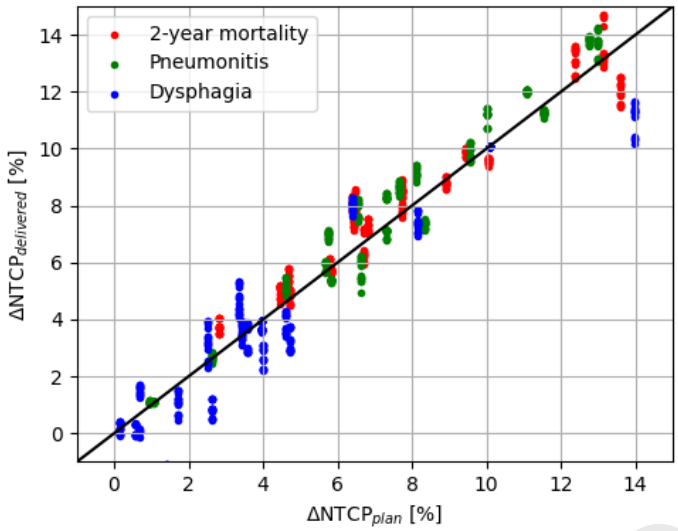


Figure 4: Scatter plot of the planned ΔNTCP versus the delivered ΔNTCP for all values of α/β and all DIR algorithms.

The planned and delivered ΔNTCP for an individual patient can nevertheless differ substantially (Figure 4), and the robustness of $\Delta\mathrm{NTCP}_{\mathrm{plan}}$ is partly because it is well above the threshold for most studied cases. Despite the variations for individual patients, the average difference between the planned and delivered ΔNTCP was found to not significantly differ from 0, i.e. $\Delta\mathrm{NTCP}_{\mathrm{plan}}$ is an unbiased predictor of $\Delta\mathrm{NTCP}_{\mathrm{delivered}}$ (Table 1). However, looking at the endpoints individually, the advantage of protons is significantly underestimated in $\Delta\mathrm{NTCP}_{\mathrm{plan}}$ for both 2-year mortality and radiation pneumonitis (Table 1). For dysphagia, the advantage is overestimated, but this is not significant (p=0.05).

The 95% confidence interval of $\Delta\mathrm{NTCP}_{\mathrm{delivered}} - \Delta\mathrm{NTCP}_{\mathrm{plan}}$ shows that the patient selection threshold may be violated if $\Delta\mathrm{NTCP}_{\mathrm{plan}}$ surpasses it mildly (Table 1). For example, if $\Delta\mathrm{NTCP}_{\mathrm{plan}}$ for 2-year mortality is only 1.02 p.p. above the 2 p.p. selection threshold, $\Delta\mathrm{NTCP}_{\mathrm{delivered}}$ would fall below the threshold in 2.5% of all cases. Generally, for all endpoints, care should be taken if the threshold is surpassed with less than 2 p.p.

Table 1: Mean difference between the planned and delivered ΔNTCP for each endpoint. A paired t-test is used to test whether the mean $\Delta NTCP_{plan}$ and $\Delta NTCP_{delivered}$ are significantly different. The 95% confidence interval of the individual difference is shown in the third column.

	Mean $\Delta\mathrm{NTCP}_{\mathrm{delivered}} - \Delta\mathrm{NTCP}_{\mathrm{plan}}$	95% confidence interval $\Delta\mathrm{NTCP}_{\mathrm{delivered}} - \Delta\mathrm{NTCP}_{\mathrm{plan}}$
All endpoints	0.07 p.p. (p = 0.15)	[-2.08, 2.22] p.p.
• 2-year mortality	0.42 p.p. (p = 0.01)	[-1.02, 1.88] p.p.
• Radiation Pneumonitis	0.40 p.p. (p = 0.02)	[-0.99, 1.80] p.p.
• Dysphagia	-0.62 p.p. (p = 0.05)	[-3.28, 2.05] p.p.

Discussion

In this work, the NTCP of the delivered proton treatment is estimated by dose reconstruction and inter- and intrafraction accumulation. $\Delta\mathrm{NTCP}_{\mathrm{delivered}}$ is calculated between the delivered proton treatment and the planned photon treatment, because the patient was not treated with photons so the log-file data to reconstruct the treatment is unavailable. This is a limitation, because

uncertainties in delivery, DIR and α/β ratio for the photon plan are not considered. However, because of the lack of interplay effect in photon therapy [32] and its lower sensitivity to density changes, the variations in NTCP between plan and delivery are expected to be smaller than for protons. Nevertheless, this assumption requires validation and is a subject of current work at the UMCG.

4DREAL incorporates the effects of weekly anatomical changes and the interplay effect. However, it does not account for systematic setup errors or irregular breathing motion [11][25][25][25][27]. Even though the 4DCT is acquired in treatment position and aligned by laser, this position can deviate from the actual treatment position. In fact, the actual treatment position is likely better aligned with the plan because of the additional gantry-mounted cone-beam CT guidance. In future work, this cone-beam CT could be used to directly acquire the daily anatomy and position right before the delivery [33]. Irregular breathing motion exhibiting varying amplitudes and baseline drifts [34] could be accounted for using advanced motion modelling approaches [35–37].

Our results show that the effect of DIR on the NTCP is larger than the effect of α/β ratio, but this conclusion is subject to two considerations. Firstly, the dose reconstruction was limited to one fraction per repeated 4DCT. Including all fractions of a treatment during accumulation would smooth out the random differences between DIRs and therefore weaken the DIR dependence [11]. Our results, however, indicate that a part of the DIR effect is systematic and that part would not cancel out. Secondly, we assumed that the repeated 4DCTs are representative of the daily variations of the whole treatment. If larger variations occurred in reality, the effect of changing the α/β ratio would increase. However, in clinical practice the patient anatomy is verified daily during the positioning with CBCT and a new 4DCT is acquired if anatomical changes are detected. Therefore, despite these considerations, we expect that the influence of DIR will be larger than that of α/β, especially because the DIR influence is approximately 10 times higher in our results.

Despite the low dose and large dose gradients in the OARs, the influence of daily dose variations on treatment outcome was found to be small. The biological effect can be locally large, but as NTCP is calculated for the mean OAR dose, the effect is moderated. However, none of the patients in this study underwent replanning, which would increases the daily dose variations. Therefore, in future work, our findings should be tested including replanned cases.

The biologically equivalent dose translates the effect of non-uniform fraction doses into NTCP. However, NTCP models are tuned on actual patient data, therefore already incorporating the effect of daily dose variations. Thus, using the biologically equivalent dose may account for this effect twice. It is however important to include this effect in case patients undergo larger dosimetric variations than the average, which could be the case for the patients in this study. Nevertheless, the effect was found to be small.

We found differences up to 2 p.p. NTCP between planned and delivered treatments, even though the differences are usually much smaller. This means that patients who surpass the threshold for patient selection with less than 2 p.p. might not have been selected based on delivered dose. However, since these thresholds are clinically arbitrary, this is not directly correlated to the benefit of proton therapy as many toxicities are not taken into account.

Conclusion

For the patient population and treatment technique employed, we found that $\Delta\mathrm{NTCP}_{\mathrm{plan}}$ is on average an unbiased predictor of $\Delta\mathrm{NTCP}_{\mathrm{delivered}}$, but that the difference can vary up to 2 p.p. for a

single case. Despite that, the clinical benefit of protons for patients selected for proton therapy was confirmed for 97.0 % percent of all cases included.

Our results further show that the effect of DIR cannot be neglected in dose accumulation, as differences in DIR algorithms result in NTCP differences up to 1.6 p.p. The influence of varying α/β was found to be one order of magnitude lower.

Acknowledgment

This project has received funding from the European Union's Horizon 2020 Marie Skłodowska-Curie Actions under Grant Agreement No. 955956. The authors would like to thank Cosylab for providing their DIR algorithm.

References

[1] Langendijk JA, Lambin P, De Ruysscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: The model-based approach. Radiother Oncol 2013;107:267–73. https://doi.org/https://doi.org/10.1016/j.radonc.2013.05.007.

[2] Palma G, Monti S, Conson M, Pacelli R, Cella L. Normal tissue complication probability (NTCP) models for modern radiation therapy. Semin Oncol 2019;46:210–8. https://doi.org/10.1053/J.SEMINONCOL.2019.07.006.

[3] Lomax AJ. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions. Phys Med Biol 2008;53:1043–56. https://doi.org/10.1088/0031-9155/53/4/015.

[4] Albertini F, Bolsi A, Lomax AJ, Rutz HP, Timmerman B, Goitein G. Sensitivity of intensity modulated proton therapy plans to changes in patient weight. Radiother Oncol 2008;86:187–94.

[5] 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. https://doi.org/10.1088/0031-9155/52/10/009.

[6] Engelsman M, Kooy HM. Target volume dose considerations in proton beam treatment planning for lung tumors. Med Phys 2005;32:3549–57. https://doi.org/https://doi.org/10.1118/1.2126187.

[7] Pedroni E, Bacher R, Blattmann H, Böhringer T, Coray A, Lomax A, et al. The 200-MeV proton therapy project at the Paul Scherrer Institute: Conceptual design and practical realization. Med Phys 1995;22:37–53. https://doi.org/https://doi.org/10.1118/1.597522.

[8] Grassberger C, Dowdell S, Lomax A, Sharp G, Shackleford J, Choi N, et al. Motion Interplay as a Function of Patient Parameters and Spot Size in Spot Scanning Proton Therapy for Lung Cancer. Int J Radiat Oncol 2013;86:380–6. https://doi.org/10.1016/J.IJROBP.2013.01.024.

[9] Bert C, Rietzel E. 4D treatment planning for scanned ion beams. Radiat Oncol 2007;2:24. https://doi.org/10.1186/1748-717X-2-24.

[10] Meijers A, Jakobi A, Stützer K, Guterres Marmitt G, Both S, Langendijk JA, et al. Log file-based dose reconstruction and accumulation for 4D adaptive pencil beam scanned proton therapy in a clinical treatment planning system: Implementation and proof-of-concept. Med Phys 2019;46:1140–9.

[11] Meijers A, Knopf A-C, Crijns APG, Ubbels JF, Niezink AGH, Langendijk JA, et al. Evaluation of

interplay and organ motion effects by means of 4D dose reconstruction and accumulation. Radiother Oncol 2020;150:268–74.

[12] Nederlandse Vereniging voor Radiotherapie en Oncologie. Landelijk Indicatie Protocol Protonen Therapie - Longcarcinoom. 2019.

[13] Ribeiro CO, Knopf A, Langendijk JA, Weber DC, Lomax AJ, Zhang Y. Assessment of dosimetric errors induced by deformable image registration methods in 4D pencil beam scanned proton treatment planning for liver tumours. Radiother Oncol 2018;128:174–81. https://doi.org/https://doi.org/10.1016/j.radonc.2018.03.001.

[14] Nenoff L, Ribeiro CO, Matter M, Hafner L, Josipovic M, Langendijk JA, et al. Deformable image registration uncertainty for inter-fractional dose accumulation of lung cancer proton therapy. Radiother Oncol 2020;147:178–85. https://doi.org/10.1016/J.RADONC.2020.04.046.

[15] Paganetti H, Niemierko A, Ancukiewicz M, Gerweck LE, Goitein M, Loeffler JS, et al. Relative biological effectiveness (RBE) values for proton beam therapy. Int J Radiat Oncol 2002;53:407–21. https://doi.org/https://doi.org/10.1016/S0360-3016(02)02754-2.

[16] McNamara AL, Schuemann J, Paganetti H. A phenomenological relative biological effectiveness (RBE) model for proton therapy based on all publishedin vitrocell survival data. Phys Med Biol 2015;60:8399–416. https://doi.org/10.1088/0031-9155/60/21/8399.

[17] Pfeiler T, Bäumer C, Engwall E, Geismar D, Spaan B, Timmermann B. Experimental validation of a 4D dose calculation routine for pencil beam scanning proton therapy. Z Med Phys 2018;28:121–33. https://doi.org/https://doi.org/10.1016/j.zemedi.2017.07.005.

[18] Batista V, Richter D, Chaudhri N, Naumann P, Herfarth K, Jäkel O. Significance of intra-fractional motion for pancreatic patients treated with charged particles. Radiat Oncol 2018;13:120. https://doi.org/10.1186/s13014-018-1060-8.

[19] Stewart JR, Fajardo LF, Gillette SM, Constine LS. Radiation injury to the heart. Int J Radiat Oncol Biol Phys 1995;31:1205–11.

[20] Gillette EL, McChesney SL, Hoopes PJ. Isoeffect curves for radiation-induced cardiomyopathy in the dog. Int J Radiat Oncol Biol Phys 1985;11:2091–7.

[21] Schultz-Hector S, Sund M, Thames HD. Fractionation response and repair kinetics of radiation-induced heart failure in the rat. Radiother Oncol 1992;23:33–40.

[22] Van Dyk J, Mah K, Keane TJ. Radiation-induced lung damage: dose-time-fractionation considerations. Radiother Oncol 1989;14:55–69.

[23] Scheenstra AEH, Rossi MMG, Belderbos JSA, Damen EMF, Lebesque J V, Sonke J-J. Alpha/Beta Ratio for Normal Lung Tissue as Estimated From Lung Cancer Patients Treated With Stereotactic Body and Conventionally Fractionated Radiation Therapy. Int J Radiat Oncol 2014;88:224–8. https://doi.org/https://doi.org/10.1016/j.ijrobp.2013.10.015.

[24] Meijers A, Seller OC, Free J, Bondesson D, Oria CS, Rabe M, et al. Assessment of range uncertainty in lung-like tissue using a porcine lung phantom and proton radiography. Phys Med Biol 2020;65:155014.

[25] Guterres Marmitt G, Pin A, Ng Wei Siang K, Janssens G, Souris K, Cohilis M, et al. Platform for automatic patient quality assurance via Monte Carlo simulations in proton therapy. Phys Medica 2020;70:49–57. https://doi.org/10.1016/J.EJMP.2019.12.018.

[26] Brock KK, Consortium DRA. Results of a multi-institution deformable registration accuracy study (MIDRAS). Int J Radiat Oncol Biol Phys 2010;76:583–96.

[27] Kadoya N, Nakajima Y, Saito M, Miyabe Y, Kurooka M, Kito S, et al. Multi-institutional Validation Study of Commercially Available Deformable Image Registration Software for Thoracic Images. Int J Radiat Oncol 2016;96:422–31. https://doi.org/https://doi.org/10.1016/j.ijrobp.2016.05.012.

[28] Bortfeld T, Jokivarsi K, Goitein M, Kung J, Jiang SB. Effects of intra-fraction motion on IMRT dose delivery: statistical analysis and simulation. Phys Med Biol 2002;47:302. https://doi.org/10.1088/0031-9155/47/13/302.

[29] Fowler JF. The linear-quadratic formula and progress in fractionated radiotherapy. Br J Radiol 1989;62:679–94.

[30] Klement RJ, Sonke J-J, Allgäuer M, Andratschke N, Appold S, Belderbos J, et al. Estimation of the α/β ratio of non-small cell lung cancer treated with stereotactic body radiotherapy. Radiother Oncol 2020;142:210–6.

[31] Wu AJ, Williams E, Modh A, Foster A, Yorke E, Rimner A, et al. Dosimetric predictors of esophageal toxicity after stereotactic body radiotherapy for central lung tumors. Radiother Oncol 2014;112:267–71. https://doi.org/https://doi.org/10.1016/j.radonc.2014.07.001.

[32] Rao M, Wu J, Cao D, Wong T, Mehta V, Shepard D, et al. Dosimetric Impact of Breathing Motion in Lung Stereotactic Body Radiotherapy Treatment Using Image-Modulated Radiotherapy and Volumetric Modulated Arc Therapy. Int J Radiat Oncol 2012;83:e251–6. https://doi.org/10.1016/J.IJROBP.2011.12.001.

[33] Thummerer A, Seller Oria C, Zaffino P, Visser S, Meijers A, Guterres Marmitt G, et al. Deep learning–based 4D-synthetic CTs from sparse-view CBCTs for dose calculations in adaptive proton therapy. Med Phys 2022;n/a. https://doi.org/https://doi.org/10.1002/mp.15930.

[34] Dhont J, Vandemeulebroucke J, Burghelea M, Poels K, Depuydt T, Van Den Begin R, et al. The long- and short-term variability of breathing induced tumor motion in lung and liver over the course of a radiotherapy treatment. Radiother Oncol 2018;126:339–46. https://doi.org/10.1016/j.radonc.2017.09.001.

[35] Steinsberger T, Alliger C, Donetti M, Krämer M, Lis M, Paz A, et al. Extension of RBE-weighted 4D particle dose calculation for non-periodic motion. Phys Medica 2021;91:62–72. https://doi.org/10.1016/J.EJMP.2021.10.009.

[36] 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.

[37] Wölfelschneider J, Seregni M, Fassi A, Ziegler M, Baroni G, Fietkau R, et al. Examination of a deformable motion model for respiratory movements and 4D dose calculations using different driving surrogates. Med Phys 2017;44:2066–76. https://doi.org/https://doi.org/10.1002/mp.12243.

Inter- and intrafractional 4D dose accumulation for evaluating ΔNTCP robustness in lung cancer: manuscript highlights

- ΔNTCP of (reconstructed) *delivered dose* is compared to ΔNTCP based on *planned dose*
- ΔNTCP based on the *plan* is an unbiased predictor of ΔNTCP based on the *delivered* dose

- For a single patient, variations of ±2% compared to delivered ΔNTCP can be expected.
- The deformable image registration algorithm affects ΔNTCP up to 1.6%.
- The choice of α/β ratio has only a minor effect on the ΔNTCP.

Title:

Inter- and intrafractional 4D dose accumulation for evaluating ΔNTCP robustness in lung cancer

Short title:

Inter- and intrafractional 4D dose accumulation for evaluating ΔNTCP robustness in lung cancer

Authors and affiliations:

Smolders Andreas [a,b], Hengeveld Adriaan C.[c], Both Stefan[c], Wijsman Robin[c], Langendijk Johannes A.[c], Weber Damien C.[a,d,e], Lomax Anthony J.[a,b], Albertini Francesca[a], Guterres Marmitt Gabriel[c]

[a] Centre for Proton Therapy, Paul Scherrer Institute, Villigen, Switzerland
[b] Department of Physics, ETH Zurich, Zurich, Switzerland
[c] Department of Radiation Oncology, University Medical Center of Groningen, University of Groningen, Groningen, The Netherlands
[d]Department of Radiation Oncology, University Hospital Zurich, Zurich, Switzerland
[e]Department of Radiation Oncology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

Corresponding author:

Andreas Smolders
Paul Scherrer Institute
WBBB/105
Forschungsstrasse 111
5232 Villigen PSI
Switzerland
+41 56 310 39 75
andreas.smolders@psi.ch

Support:

This project has received funding from the European Union's Horizon 2020 Marie Skłodowska-Curie Actions under Grant Agreement No. 955956.

Inter- and intrafractional 4D dose accumulation for evaluating ΔNTCP robustness in lung cancer: conflict of interest statement

The authors have no conflict of interest to declare.

Andreas Smolders, on behalf of the authors