IL4I1 Is a Metabolic Immune Checkpoint that Activates the AHR and Promotes Tumor Progression

Sadik, Ahmed; Somarribas Patterson, Luis F; Öztürk, Selcen; Mohapatra, Soumya R; Panitz, Verena; Secker, Philipp F; Pfänder, Pauline; Loth, Stefanie; Salem, Heba; Prentzell, Mirja Tamara

Published in:
Cell

DOI:
10.1016/j.cell.2020.07.038

Publication date:
2020

Journal Pre-proofs

Original Article

Head and Neck IMPT probabilistic dose accumulation: feasibility of a 2 mm setup uncertainty setting

Dirk Wagenaar, Roel G.J. Kierkels, Arjen van der Schaaf, Arturs Meijers, Daniel Scandurra, Nanna M. Sijtsema, Erik W. Korevaar, Roel J.H.M. Steenbakkers, Antje C. Knopf, Johannes A. Langendijk, Stefan Both

PII: S0167-8140(20)30770-2
DOI: https://doi.org/10.1016/j.radonc.2020.09.001
Reference: RADION 8502

To appear in: *Radiotherapy and Oncology*

Received Date: 14 February 2020
Revised Date: 14 August 2020
Accepted Date: 2 September 2020

Please cite this article as: Wagenaar, D., Kierkels, R.G.J., van der Schaaf, A., Meijers, A., Scandurra, D., Sijtsema, N.M., Korevaar, E.W., Steenbakkers, R.J.H., Knopf, A.C., Langendijk, J.A., Both, S., Head and Neck IMPT probabilistic dose accumulation: feasibility of a 2 mm setup uncertainty setting, *Radiotherapy and Oncology* (2020), doi: https://doi.org/10.1016/j.radonc.2020.09.001

Head and Neck IMPT probabilistic dose accumulation: feasibility of a 2 mm setup uncertainty setting.

Authors

1. Dirk Wagenaar[1] (MSc)
2. Roel G.J. Kierkels[1, 2] (PhD)
3. Arjen van der Schaaf[1] (PhD)
4. Arturs Meijers[1] (MSc)
5. Daniel Scandurra[1] (MSc)
6. Nanna M. Sijtsema[1] (PhD)
7. Erik W. Korevaar[1] (PhD)
8. Roel J.H.M. Steenbakkers[1] (PhD)
9. Antje C. Knopf[1] (PhD)
10. Johannes A. Langendijk[1] (MD/PhD)
11. Stefan Both[1] (PhD)

Affiliations

1. Department of Radiation Oncology,
   University Medical Center Groningen,
   University of Groningen, The Netherlands
2. Department of Radiation Oncology,
   Radiotherapiegroep, Arnhem/Deventer,
   The Netherlands

Corresponding author's details:

Dirk Wagenaar, P.O. Box 30.001, 9700 RB Groningen, the Netherlands
ir.d.wagenaar@umcg.nl
0652724467

Financial disclosure:

Nothing to disclose

Running head:

Head and Neck IMPT: a 2 mm setup uncertainty

ABSTRACT

(MAX. 250 WORDS)

Objective

To establish optimal robust optimization uncertainty settings for clinical head and neck cancer (HNC) patients undergoing 3D image-guided pencil beam scanning (PBS) proton therapy.

Methods

We analyzed ten consecutive HNC patients treated with 70 and 54.25 Gy_{RBE} to the primary and prophylactic clinical target volumes (CTV) respectively using intensity-modulated proton therapy (IMPT). Clinical plans were generated using robust optimization with 5 mm/3% setup/range uncertainties (RayStation v6.1). Additional plans were created for 4, 3, 2 and 1 mm setup and 3% range uncertainty and for 3 mm setup and 3%, 2% and 1% range uncertainty.

Systematic and random error distributions were determined for setup and range uncertainties based on our quality assurance program. From these, 25 treatment scenarios were sampled for each plan, each consisting of a systematic setup and range error and daily random setup errors. Fraction doses were calculated on the weekly verification CT closest to the date of treatment as this was considered representative of the daily patient anatomy.

Results

Plans with a 2 mm/3% setup/range uncertainty setting adequately covered the primary and prophylactic CTV ($V_{95} \geq 99\%$ in 98.8% and 90.8% of the treatment scenarios respectively). The average organ-at-risk dose decreased with 1.1 Gy_{RBE}/mm setup uncertainty reduction and 0.5 Gy_{RBE}/1% range uncertainty reduction. Normal tissue complication probabilities decreased by 2.0%/mm setup uncertainty reduction and by 0.9%/1% range uncertainty reduction.

Conclusion

The results of this study indicate that margin reduction below 3 mm/3% is possible but requires a larger cohort to substantiate clinical introduction.

INTRODUCTION

The goal of radiotherapy treatment plan robust optimization is to create a deliverable treatment plan which adequately covers the target volume with the lowest dose possible to the most relevant organs at risk. Treatment preparation and execution uncertainties need to be accounted for to avoid undertreating the target.[1] Composite minimax robust optimization (CMRO) is known to be advantageous for pencil beam scanning (PBS) proton therapy as compared to PTV-based treatment planning which is more sensitive to rigid shifts.[2] Conventionally, the isocentric uncertainty setting used in CMRO is determined by margin recipes for X-ray radiotherapy.[3] For example, van Herk calculated the required CTV-PTV margin to treat the clinical target volume (CTV) to 95% of the prescribed dose for 90% of the patients based on the systematic and random isocentric uncertainty.[1,3] However, the risk of undertreating the target should be weighed against the risk of toxicity to find the optimal uncertainty setting.[4] Furthermore, the van Herk formula assumes dose invariance to external and internal movement which does not hold for proton therapy.[3,5–7]

Van der Water et al. studied the effect of setup and range uncertainty setting reduction and found a gradual decrease in normal tissue complication probabilities (NTCPs) as uncertainties were reduced.[4] However, their study did not take anatomical changes or treatment uncertainties into consideration as the target is always covered in the nominal scenario.[4] Changes in anatomy can have a large impact on the delivered dose distribution compared to the planned dose distribution.[8–10] These and other results show that uncertainty setting reduction is impactful on patient-reported toxicities. Previous studies in head and neck cancer (HNC) photon therapy reported toxicity reduction after changing the CTV-planning target volume (PTV) margin from 5 mm to 3 mm, but a further reduction to a 2 mm uncertainty setting might be possible without altering the physician prescribed dosimetric parameters.[11,12] By accumulating the dose on diagnostic quality verification CT images, the impact of interfractional anatomy changes can be evaluated and the delivered dose can be estimated more accurately.[8–10,13] Earlier studies focused on dose accumulation incorporating repeated imaging to study the effect of changing anatomy, but these neglected other treatment uncertainties such as inter- and intrafraction motion which could have a large effect on target coverage.[8,9]

In addition to anatomical changes, positional and treatment uncertainties have a large effect on the delivered proton therapy dose distribution in and near the target.[5–7] If the probability density distributions of systematic and random positioning and range errors are known, the probability density distribution of dosimetric parameters can be calculated to predict the tumor control probability (TCP) and normal tissue complication probability (NTCP).[14–16] Such approaches have been extensively studied for probabilistic treatment planning as a way to incorporate uncertainties.[6,17,18] In this study we use a probabilistic approach to retrospectively estimate the dose in different treatment scenarios so that a representative estimation of the actually given dose can be determined.

The aim of this study is to establish whether CMRO uncertainty settings can be reduced below 3 mm for HNC IMPT treatments using probabilistic dose accumulation. The probabilistic dose accumulation [illegible]ent treatment scenarios incorporating the systematic and random setup and

MATERIALS AND METHODS

Patients and treatment

The study population was composed of the first ten consecutive HNC patients treated with PBS intensity-modulated proton therapy (IMPT) at our institute. Patient characteristics are further described in Table 1. In the Netherlands, patients are selected for proton therapy using model-based selection. In this approach, proton therapy is applied if the ΔNTCP between the proton and photon treatment plans exceeds a certain indication-specific threshold.[14] The ΔNTCP thresholds used for HNC patients are 10% for a single grade II complication, 15% for the combined total of two grade II complications, or 5% for a single grade III complication.[14]

Patients were treated with 70 Gy_{RBE} to the primary CTV and 54.25 Gy_{RBE} to the prophylactic CTV in 35 fractions with 5 fractions per week using a constant relative biological effectiveness (RBE) factor of 1.1. Patients were immobilized using a 5-point mask (HP Pro, Orfit Industries, Wijnegem, Belgium). For each fraction, positioning correction vectors were determined by matching daily cone-beam CT (CBCT) images to the planning CT and applied using a 6-D robotic table capable of yaw, pitch and roll corrections. IMPT treatments were delivered at our proton treatment facility (Proteus ®Plus, IBA, Ottignies-Louvain-la-Neuve, Belgium). Changes in anatomy were monitored with weekly offline verification CTs (Somatom AS Open, Siemens, München, Germany) with the patient immobilized in treatment position.

Treatment planning

Clinical treatment plans were generated using CMRO with a 3% range and 5 mm setup uncertainty in the treatment planning system (TPS) (Raystation v6.1, RaySearch Laboratories, Stockholm, Sweden). During optimization, 7 different isocenter shifts (i.e. no shift or a positive or negative shift along one of the three axes) and two (i.e. positive or negative) density shifts are considered. The worst target dose in these 14 scenarios is optimized.

Target coverage was assessed using the voxel-wise minimum robustness (multi-scenario) evaluation approach outlined in a recent publication describing the Dutch consensus for proton plan evaluation.[19] During evaluation, 14 different isocenter shifts (i.e. a positive or negative shift along each of the three axes and along each diagonal) and two density shits were considered resulting in a total of 28 scenario's.[19] The voxel-wise minimum of these 28 scenarios was used to assess robust target coverage, where the coverage criteria was a V_{95} of both CTVs of at least 98%. Eight patients were treated with a four-beam setup with gantry angles between 40-60, 160, 195-200 and 295-320 degrees. For one patient, treated for left-sided retromolar trigonum squamous cell carcinoma, the right-sided posterior beam was omitted. Another patient, treated for left-sided tonsillar carcinoma, was treated with unilateral irradiation with three beams angled at 45, 90 and 135 degrees.

Treatment plans with various setup and range uncertainty settings are required to conduct this study. These treatment plans with various uncertainty settings were derived from the clinical plan in an automatic process. Therefore in the text we refer to these plans as adjusted treatment plans. The t plans were generated in a two-step process (Figure 1). First the dose distribution ler setup and range uncertainty is predicted. Second, this predicted dose erted into a deliverable adjusted treatment plan using a voxel-based dose

mimicking optimization approach. [20,21] This process ensures consistent treatment plan quality in terms of prioritization of the organs at risk to spare, while keeping the physician prescribed tumor coverage unaltered.

Additionally, one planner-generated treatment plans per patient was created to validate that the automatic adjustment of treatment plans gives similar results as conventionally created treatment plans. The planner-generated treatment plans were created with a 3 mm setup and 3% range uncertainty. These planner-generated treatment plans were compared to the adjusted treatment plans with a 3 mm setup and 3% range uncertainty in terms of the dose given to at least 98% of the target (D_{98}) in the voxel-wise minimum dose distribution and the dose-fall off outside the CTV in the planned scenario.

Using the automated approach, we generated seven adjusted treatment plans per patient with various setup and range uncertainty settings and one planner-generated treatment plan per patient.

Error distributions

The distributions of the intrafraction shifts were tested for normality by visually inspecting their quantile-quantile (Q-Q) plots and tested for a systematic component using a two-sided Student t-test with an α of 0.05. All other considered error distributions were assumed to be Gaussian. The accuracy of the onboard CBCT imaging system isocenter and 6-D robotic table shifts were assessed based on the results of our comprehensive machine QA program.

To estimate the intrafraction isocentric displacement, CBCTs before and after treatment were analyzed. For these patients, CBCTs were made before and after treatment for 41 fractions in total (i.e. 4.1 on average per patient) as part of the post-treatment position verification. The post-fraction CBCTs were matched using rigid registration to estimate the intrafraction isocentric displacement in three dimensions. The distribution of intrafraction displacements was taken as the random setup error distribution and statistically tested for a systematic component. The onboard CBCT and 6-D robotic couch error were quadratically summed and taken as the systematic setup error with an additional 0.5 mm (one standard deviation (SD)) in all directions to account for potentially neglected errors. The magnitude of the residual error was a conservative estimation based on our expectation of unconsidered systematic errors such as intrafraction rotations and methodological uncertainties such as deformable image registration errors.

Range errors are systematic in nature and occur depending on tissue type. Common clinical practice is to account for a 2.4% + 1.0 mm range uncertainty error during treatment optimization.[22,23] This recipe was shown to be equal to two SDs in a recent study analyzing the residual range errors in proton radiography validation of our CT calibration curve.[23] The SD of 1.2% + 0.5 mm was converted to a patient-specific percentage by dividing the absolute component by the monitor unit weighted average range of the clinical treatment plan.

Probabilistic dose accumulation

The error distributions (i.e. CBCT isocenter, robotic table shifts, intrafraction motion, range uncertainty and a residual error) were applied to weekly verification images to simulate the ion of dose distributions that would be delivered to the patient when treated eatment plans. Similar to a previous study, we define a single treatment scenario

as the accumulated dose of a complete treatment where the random errors have been sampled from their probability density distributions for each fraction.[24] For each treatment plan, 25 treatment scenarios were calculated. The workflow of a single treatment scenario calculation is illustrated in figure 2. Seven patients had seven verification CTs and three patients had six verification CTs available. One systematic setup (originating from the CBCT isocenter, robotic table shifts and a residual error) and one systematic range error (originating from proton radiography measurements) were sampled from the error distributions described in the previous subsection. Daily fraction doses were calculated on the weekly verification CT closest to the date of treatment as this was considered to be representative of the patient anatomy of that fraction. In addition to the systematic errors, daily fraction dose calculations included a random setup error (originating from the intrafraction motion). The weekly verification CTs were mapped to the planning CT using a deformable image registration technique previously described by Weistrand et al..[25] Using this, the daily fraction doses were mapped to the planning CT and summed, resulting in an estimated delivered dose distribution for that treatment scenario calculation.

The sampled errors were different for each treatment scenario calculation but identical between adjusted treatment plans. A single treatment scenario calculation consisted of 35 fraction dose calculations, each treatment plan was simulated 25 times, each patient had 7 adjusted treatment plans with different uncertainty optimization settings resulting in 61.250 fractional dose calculations in total. Doses were calculated within a 1.0% statistical uncertainty with 3x3x3 mm^3 dose voxels using the Monte Carlo dose engine integrated in the TPS.

Statistical analysis

In our clinical practice, a target coverage criterion of V_{95} of at least 98% in the voxel-wise minimum dose distribution is applied during treatment. [19] The required coverage of the delivered dose distribution should be more than 98% as the coverage in the voxel-wise minimum dose distribution is less favorable than any single scenario, but less than 100% as this would include clinically irrelevant volumes. When evaluating the delivered dose, a 90% pass-rate is typically accepted since accounting for all possible error scenarios would result in overly large uncertainty settings.[3] In this study, we defined the delivered dose criterion as $V_{95} \geq 99\%$ which should be met for at least 90% of the treated patient population.

The target dose was evaluated in terms of the average V_{95} and the fraction of scenarios which met the coverage criteria of $V_{95} \geq 99\%$ for the primary and prophylactic CTVs. Target coverage was further reported in terms of the average tumor control probability (TCP), calculated using a TCP model by Luhr et al..[16] Normal tissue dose was evaluated in terms of the mean dose to OARs and the average NTCP values for xerostomia, grade 2-4 dysphagia, and tube feeding dependence.[14,15]The TCP is calculated based on the DVH in the three regions of the target (i.e. the gross tumor volume (GTV), the primary CTV excluding the GTV and the prophylactic CTV excluding the primary CTV). The sensitivity of the TCP to underdosage of each target region is based on the rate of recurrences in that target region. We chose the other parameters used in the TCP calculation identical to the estimations in the model publication for HNC, which were a tumor control dose D_{50} of 70 Gy_{RBE} and slope γ_{50} of 1.5.[16] The proportion of recurrences was taken from the model which was based on a study by Due et al who found that the recurrences in volumes were 82% for the GTV, 16% for the remainder of the primary CTV and of the prophylactic CTV.[16,26] Due to the low number of recurrences in the

prophylactic CTV, the TCP model is less sensitive to underdosage of the prophylactic CTV. An additional analysis is performed, also using the TCP model by Luhr et al., but using the recurrence rates of 51.3%, 29.4% and 19.3% for the GTV, the primary CTV and the prophylactic CTV respectively based on the proportion of recurrences reported to occur in these structures after IMRT in three centers from a recent study. [27]

Insufficient data is available to make an accurate substantiated estimate of D_{50} which has a large impact on the calculated TCP. Therefore, the calculated TCP does not reflect our clinical results but can be taken as a relative measure to compare the expected effectiveness of different target dose distributions. At these settings, a homogenous dose of 70 Gy_{RBE} to the both CTVs yields a TCP of 50% and a 0.4 Gy_{RBE} dose reduction results in a 1.0% point TCP reduction.

The averages of target V_{95}, mean OAR dose, NTCP and TCP were calculated for all patients and simulations which were then tested for statistical significance by calculating two-tailed p-values using linear regression analysis (R v3.5.1, R Foundation, Vienna, Austria). The statistical significance was determined after accounting for multiple testing using Bonferroni's correction.[28] Differences were considered statistically significant if $p < 0.025$ ($\alpha = 0.05/2$ parameters) for target dose, $p < 0.0055$ ($\alpha = 0.05/11$ parameters) for OAR dose, $p < 0.013$ ($\alpha = 0.05/4$ parameters) for NTCPs and $p < 0.050$ ($\alpha = 0.05/1$ parameter) for TCP.

RESULTS

The adjusted treatment plans with a 3-mm setup uncertainty created using dose mimicking optimization resulted in a slightly higher D_{98} of the primary CTV (from 67.0 to 67.1 Gy_{RBE}, $p = 0.82$) and the prophylactic CTV (from 52.1 to 52.7 Gy_{RBE}, $p = 0.19$) in the voxel-wise minimum dose distribution compared to the planner-generated treatment plans with identical uncertainty settings (figure S1). On average, the OAR dose was 0.1 Gy_{RBE} lower for the adjusted treatment plans with smaller uncertainty settings, but this was not statistically significant ($p = 0.85$) (figure S1).

The systematic component of the intrafraction motion derived from the pre and post-treatment CBCTs was not statistically significant ($p = 0.57$, 0.46 and 0.69 for lateral, longitudinal and height respectively) (figure S2). The intrafraction motion was therefore considered as random with a 0.7 mm SD in each direction. The error of the onboard CBCT imaging system isocenter and 6-D robotic couch translation were found to be 0.4 and 0.3 mm (1 SD) respectively. This, together with the 0.5 mm residual error resulted in a 0.7 mm systematic setup error and a 0.7 mm random setup error at the isocenter (one SD, all directions). The patient-specific range error distribution SD ranged from 1.6% to 1.8%.

The results of the treatment scenarios for all adjusted treatment plans are both described below and shown in table 2.

Target coverage and TCP values reduced with smaller setup and range uncertainties (figure 3). Adjusted treatment plans with a 5, 4, 3 and 2mm setup and 3% range uncertainties met the target [illegible] (90% of simulations $V_{95} \geq 99\%$) for both CTVs. When reducing the setup uncertainty [illegible] 1 mm the fraction of simulations meeting the target coverage criteria was [illegible] to 74.8% for the primary CTV and from 90.8% to 70.8% for the prophylactic

CTV. The lower dose to the CTVs resulted in a TCP reduction of 0.2%/mm setup uncertainty (p=0.013) and 0.1%/% range uncertainty (not significant). Changing the TCP model to use the recurrence rates found by another study resulted in the same values for TCP reduction as a function of setup and range uncertainty setting (Figure S3). [27]

The average dose to all OARs was reduced linearly by 1.1 Gy_{RBE} /mm setup uncertainty (min = 0.6 Gy_{RBE}/mm, max = 1.4 Gy_{RBE}/mm) and 0.5 Gy_{RBE} /% range uncertainty (min = 0.1 Gy_{RBE}/%, max = 1.3 Gy_{RBE}/%) (figure 4). The dose differences translated into an NTCP reduction of 1.0%/mm (p<0.001), 0.7%/mm (p<0.001) and 0.2%/mm (p<0.001) setup uncertainty and 0.4%/% (not significant), 0.5%/% (not significant)and 0.1%/% (p<0.001) range uncertainty for xerostomia, grade 2-4 dysphagia and tube feeding dependence respectively.

DISCUSSION

In this study, we found a $V_{95} \geq 99\%$ for the primary and prophylactic CTVs for over 90% of the treatment scenarios for a 2 mm setup uncertainty or larger. We found a small decrease of target dose which resulted in a TCP reduction of 0.2%/mm setup uncertainty and 0.1%/% range uncertainty. The V_{95} of all adjusted treatment plans was lower for the prophylactic CTV than for the primary CTV which may be caused by the larger size and anatomical location of the prophylactic CTV, making its coverage more sensitive to rotations and anatomical changes. As the recurrence rates of the GTV and primary CTV are higher compared to that of the prophylactic CTV, decreased coverage of the prophylactic CTV may be clinically less relevant.[16,26] Ideally, reduction of the setup or range uncertainty setting of the robust optimization is performed by weighing the benefit in toxicity against the cost in tumor control.

In our study, we only found a marginal effect on TCP calculated according to Luhr et al. even when lowering the setup uncertainty to 1 mm during treatment planning optimization.[16] The TCP is calculated based on the DVH in the three regions of the target (i.e. the GTV, the primary CTV excluding the GTV and the prophylactic CTV excluding the primary CTV). The impact of underdosage of different regions is based on the rate of recurrences in those regions; most recurrences occur in the GTV. As a result, the TCP model is most sensitive to underdosage of the GTV and therefore, the TCP is not severely impacted by reducing the setup or range uncertainty as underdosage is most likely to occur in the GTV-CTV margin and the prophylactic CTV (table 2). The TCP model also depends on the proportion of recurrences occurring in the different sub-volumes. Therefore we redid the analysis with the same TCP model but with the proportions of recurrences found in a recent study, but this change did not impact our results. [27] While the found differences in TCP were small, the model was calibrated with the conservative assumption that the TCP is 50% for a homogeneous dose of 70 Gy_{RBE} to the CTV. In practice, the TCP values vary based on patient characteristics such as p16/HPV positivity which would result in a flatter TCP slope and therefore even smaller TCP differences.[29]

A minor increase in the TCP of the GTV at the last decimal level was observed when the setup uncertainty decreases from 2mm to 1 mm (Table 2). The magnitude of change is trivial and can be caused by a random component during Monte Carlo optimization and calculation. Nevertheless, the [illegible] for these setup uncertainty settings and this variation has no impact on our

While TCP may be related more directly to local control, DVH parameters are used to asses target coverage in our clinical practice. In our study, we considered a V_{95} of 99% adequate for coverage of the CTV for 90% of the treatment scenarios. However, no clear coverage criterion exists for the actually given dose to the CTV. A V_{95} of 98% is typically aimed for when evaluating the nominal dose to the PTV.[1] However, dose fractionation will cause dose gradients to be less steep (i.e. "blurring"), which washes out some under- and overdosage.[30] Therefore, we chose to apply a 99% coverage as a more stringent criterion to CTV coverage in the estimated actually given dose distribution. If a 98% coverage criteria would have been applied, a setup uncertainty as low as 1 mm would pass for 90% of the treatment scenarios (Figure S4).

The average OAR dose decreased by 1.1 Gy_{RBE} /mm setup and 0.5 Gy_{RBE} /% range uncertainty resulting in a summed NTCP reduction of 2.0%/mm setup and 0.9%/% range uncertainty. The relation with the range uncertainty setting was not shown to be statistically significant for all tested OARs and NTCP models. The impact on NTCP per 1 mm setup uncertainty reduction was about equal to that of 2% range uncertainty reduction. These results indicate that the setup uncertainty has a higher impact on non-target dose and toxicities than the range uncertainty, indicating the reducing setup uncertainty is of a higher priority than range uncertainty. Currently, we are investigating this further based on daily volumetric imaging and proton radiography for HNC patients treated with PBS in our clinic.

The impact of robustness parameters on NTCP values was previously investigated in a planning comparison study in 20 oropharyngeal cases treated with IMPT therapy by van de Water et al.[4] This study found a 0.59%/mm (unilateral) and 1.82%/mm (bilateral) setup uncertainty reduction for dysphagia and a 0.11%/% (unilateral) and 0.06%/% (bilateral) range uncertainty reduction for dysphagia. These results are similar to the average dysphagia probability reduction found in this study of 1.0%/mm setup uncertainty and 0.1%/% range uncertainty. Navran et al. investigated the clinical impact of reducing the CTV-PTV margin retrospectively after reducing their clinical PTV margin for X-ray therapy from 5 mm to 3 mm and found a reduction in the rate of grade 2 xerostomia (4.2%), grade ≥2 dysphagia (9.4%) and tube feeding dependence (11.4%).[11] These differences are notably higher than the calculated reductions from our study which could be partially explained by the fact that the use of retrospective data does not rule out contributions of other time-dependent factors to the decrease in toxicity.[11] Another influencing factor could be that the patients from the study by Navran et al. were treated with X-ray therapy where the PTV margin may have a larger impact on toxicity due to the larger low and medium dose-bath.[11,31]

Our study has some limitations as not all errors could be fully integrated into the treatment scenario calculation. First, the analysis was done entirely on verification CT images but positioning is typically better at the treatment room because of the use of various positioning verification systems leading to an overestimation of the variability in treatment positioning. Second, including more patients would be beneficial as it would include a wider range of anatomical changes but would require a substantial time investment. This was partly alleviated by evaluating multiple treatment scenarios on ten patients, which allowed 250 different treatment scenarios to be calculated. Third, treatment plan adaptations were not taken into consideration and could lead to better target coverage for lower margins as it prevents underdosage by compensating for anatomical changes. Fourth, the was determined only at the isocenter, intrafraction motion may have had a nt if more scans were analyzed per patient, while deformable image registration

errors between the planning and verification CTs may have occurred. Lastly, earlier studies showed that the average deformation error is in the order of a millimeter, but this deformation error was present for all adjusted treatment plans and therefore does not bias the results.[25,32]

The isocentric intrafraction motion found in our study was normally distributed with a standard deviation of 0.7 mm in each direction. This is relatively small compared to a recent study by Bruijnen et al. investigating the maximum tumor motion using MR in 84 patients who found the 95th percentile of the intrafraction motion to be up to 2.4 mm.[33] In our study, intrafraction motion was only recorded at the isocenter potentially leading to different results. In the study by Bruijnen et al., larynx patients had the largest intrafraction motion which could also explain the differences in our results as no larynx patients were included in this study.[33]

Several measures could improve the treatment accuracy which could further improve these results. A large part of the systematic setup error was the 0.5 mm standard deviation in each direction to account for potentially neglected errors. These errors include delineation errors, registration accuracy and intrafraction deformations. The delineation error was already kept to a minimum as at our institution the target delineation for every patient is evaluated by the entire HNC team which includes not only the HNC radiation oncologists but also a HNC radiologist. In this way, inter-physician delineation variability is accounted for in the target delineation process itself. By improving deformable image registration accuracy and the intrafraction motion assessment, the residual error used in the analysis can be reduced, resulting in smaller shifts and better coverage. Moreover, as more experience is gained, the interfractional motion is expected to decrease due to improved immobilization and treatments are expected to become more robust to motion due to improvements in treatment planning and plan adaptations. Lastly, optimizing patient positioning to minimize the dose perturbation can improve target coverage without increasing the setup uncertainty setting.[34] Future work on estimation of stopping power ratios using dual-energy CT can help reduce the range errors found in proton radiography and reduce the required range uncertainty setting.[23,35] Additionally, future work should be extended to photon therapy where both robust optimization and a probabilistic assessment of target coverage should be used to further improve treatment.[36]

This is the first study to incorporate anatomical changes and systematic and random setup and range uncertainties to determine the estimated delivered dose by means of dose mapping and accumulation using longitudinal CT imaging in IMPT HNC patients. All errors were determined for the clinical workflow for the patients under investigation such as intrafraction motion assessments, and the comprehensive machine quality assurance results of the period of treatment. Note that the use of automated adjusted treatment plans is not required for clinical implementation as planner-generated treatment plans were of similar quality. The methodology used in this study offers a more patient outcome driven approach compared to only considering the risk of missing the target as the results of different margins were evaluated in terms of clinically relevant endpoints such as TCP and NTCP.

The results of this study indicate a 2 mm/3% setup and range uncertainty is sufficient for optimizing oropharynx HNC robust treatment plans when a 5-point immobilization mask and 6D couch are [illegible]ction with daily CBCT patient alignment. While a larger cohort is needed for

clinical adoption, the results indicate that future work will be able to substantiate a 2 mm setup uncertainty setting for HNC IMPT.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Herman Credoe for creating the planner-generated treatment plans to validate the fully automated dose mimicking optimization tool.

REFERENCES

[1] ICRU. ICRU 83 Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT). vol. 10. 2010. https://doi.org/10.1093/jicru/ndq025.

[2] Fredriksson A, Bokrantz R. A critical evaluation of worst case optimization methods for robust intensity-modulated proton therapy planning. Med Phys 2014;41:081701. https://doi.org/10.1118/1.4883837.

[3] van Herk M, Remeijer P, Rasch C, Lebesque J V. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. Int J Radiat Oncol 2000;47:1121–35. https://doi.org/10.1016/S0360-3016(00)00518-6.

[4] van de Water S, van Dam I, Schaart DR, Al-Mamgani A, Heijmen BJM, Hoogeman MS. The price of robustness; impact of worst-case optimization on organ-at-risk dose and complication probability in intensity-modulated proton therapy for oropharyngeal cancer patients. Radiother Oncol 2016;120:56–62. https://doi.org/10.1016/j.radonc.2016.04.038.

[5] 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.

[6] Unkelbach J, Bortfeld T, Martin BC, Soukup M. Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning. Med Phys 2009;36:149–63. https://doi.org/10.1118/1.3021139.

[7] van der Voort S, van de Water S, Perkó Z, Heijmen B, Lathouwers D, Hoogeman M. Robustness Recipes for Minimax Robust Optimization in Intensity Modulated Proton Therapy for Oropharyngeal Cancer Patients. ` 2016;95:163–70. https://doi.org/10.1016/j.ijrobp.2016.02.035.

[8] Houweling AC, Crama K, Visser J, Fukata K, Rasch CRN, Ohno T, et al. Comparing the dosimetric impact of interfractional anatomical changes in photon, proton and carbon ion radiotherapy for pancreatic cancer patients. Phys Med Biol 2017;62:3051–64. https://doi.org/10.1088/1361-6560/aa6419.

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

[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. https://doi.org/10.1002/mp.13371.

[11] Navran A, Heemsbergen W, Janssen T, Hamming-Vrieze O, Jonker M, Zuur C, et al. The impact of margin reduction on outcome and toxicity in head and neck cancer patients treated with image-guided volumetric modulated arc therapy (VMAT). Radiother Oncol 2019;130:25–31. https://doi.org/10.1016/j.radonc.2018.06.032.

[12] Chen AM, Farwell DG, Luu Q, Donald PJ, Perks J, Purdy JA. Evaluation of the Planning Target Volume in the Treatment of Head and Neck Cancer With Intensity-Modulated Radiotherapy: What Is the Appropriate Expansion Margin in the Setting of Daily Image Guidance? Int J Radiat Oncol 2011;81:943–9. https://doi.org/10.1016/j.ijrobp.2010.07.017.

[13] ndsay PE, Brock KK, Deasy JO, Tomé WA. Accurate accumulation of dose for derstanding of radiation effects in normal tissue. Int J Radiat Oncol Biol Phys

2010;76:S135-9. https://doi.org/10.1016/j.ijrobp.2009.06.093.

[14] Landelijk Platform Protonentherapie, Protonentherapie LP. Landelijk Indicatie Protocol Protonen Therapie Hoofd-halstumoren. 2017.

[15] Wopken K, Bijl HP, van der Schaaf A, van der Laan HP, Chouvalova O, Steenbakkers RJHM, et al. Development of a multivariable normal tissue complication probability (NTCP) model for tube feeding dependence after curative radiotherapy/chemo-radiotherapy in head and neck cancer. Radiother Oncol 2014;113:95–101. https://doi.org/10.1016/j.radonc.2014.09.013.

[16] Lühr A, Löck S, Jakobi A, Stützer K, Bandurska-Luque A, Vogelius IR, et al. Modeling tumor control probability for spatially inhomogeneous risk of failure based on clinical outcome data. Z Med Phys 2017;27:285–99. https://doi.org/10.1016/j.zemedi.2017.06.003.

[17] Bohoslavsky R, Witte MG, Janssen TM, van Herk M. Probabilistic objective functions for margin-less IMRT planning. Phys Med Biol 2013;58:3563–80. https://doi.org/10.1088/0031-9155/58/11/3563.

[18] Bangert M, Hennig P, Oelfke U. Analytical probabilistic modeling for radiation therapy treatment planning. Phys Med Biol 2013;58:5401–19. https://doi.org/10.1088/0031-9155/58/16/5401.

[19] Korevaar EW, Habraken SJM, Scandurra D, Kierkels RGJ, Unipan M, Eenink MGC, et al. Practical robustness evaluation in radiotherapy - A photon and proton-proof alternative to PTV-based plan evaluation. Radiother Oncol 2019. https://doi.org/10.1016/j.radonc.2019.08.005.

[20] Kierkels RGJ, Fredriksson A, Both S, Langendijk JA, Scandurra D, Korevaar EW. Automated robust proton planning using dose-volume histogram-based mimicking of the photon reference dose and reducing organ at risk dose optimization. Int J Radiat Oncol Biol Phys 2018;103:251–8. https://doi.org/10.1016/j.ijrobp.2018.08.023.

[21] Fredriksson A. Automated improvement of radiation therapy treatment plans by optimization under reference dose constraints. Phys Med Biol 2012;57:7799–811. https://doi.org/10.1088/0031-9155/57/23/7799.

[22] 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/10.1016/S0360-3016(02)02754-2.

[23] Meijers A, Free J, Wagenaar D, Deffet S, Knopf AC, Langendijk JA, et al. Validation of the proton range accuracy and optimization of CT calibration curves utilizing range probing. Phys Med Biol 2020;65:03NT02. https://doi.org/10.1088/1361-6560/ab66e1.

[24] Tilly D, Ahnesjö A. Fast dose algorithm for generation of dose coverage probability for robustness analysis of fractionated radiotherapy. Phys Med Biol 2015;60:5439–54. https://doi.org/10.1088/0031-9155/60/14/5439.

[25] Weistrand O, Svensson S. The ANACONDA algorithm for deformable image registration in radiotherapy. Med Phys 2015;42:40–53. https://doi.org/10.1118/1.4894702.

[26] Due AK, Vogelius IR, Aznar MC, Bentzen SM, Berthelsen AK, Korreman SS, et al. Recurrences after intensity modulated radiotherapy for head and neck squamous cell carcinoma more likely to originate from regions with high baseline [18F]-FDG uptake. Radiother Oncol 2014;111:360–5. https://doi.org/10.1016/j.radonc.2014.06.001.

[27] Zukauskaite R, Hansen CR, Grau C, Samsøe E, Johansen J, Petersen JBB, et al. Local [illegible]fter curative IMRT for HNSCC: Effect of different GTV to high-dose CTV margins. [illegible]col 2018;126:48–55. https://doi.org/10.1016/j.radonc.2017.11.024.

[28] Bonferroni CE. Statistical theory of classes and probability calculus. Publ R High Inst Econ Commer Sci Florence 1936;8:3–62.

[29] Psyrri A, Rampias T, Vermorken JB. The current and future impact of human papillomavirus on treatment of squamous cell carcinoma of the head and neck. Ann Oncol 2014;25:2101–15. https://doi.org/10.1093/annonc/mdu265.

[30] van Herk M, Witte M, van der Geer J, Schneider C, Lebesque J V. Biologic and physical fractionation effects of random geometric errors. Int J Radiat Oncol Biol Phys 2003;57:1460–71. https://doi.org/10.1016/j.ijrobp.2003.08.026.

[31] Arts T, Breedveld S, de Jong MA, Astreinidou E, Tans L, Keskin-Cambay F, et al. The impact of treatment accuracy on proton therapy patient selection for oropharyngeal cancer patients. Radiother Oncol 2017;125:520–5. https://doi.org/10.1016/j.radonc.2017.09.028.

[32] Kierkels RGJ, den Otter LA, Korevaar EW, Langendijk JA, van der Schaaf A, Knopf AC, et al. An automated, quantitative, and case-specific evaluation of deformable image registration in computed tomography images. Phys Med Biol 2018;63:045026. https://doi.org/10.1088/1361-6560/aa9dc2.

[33] Bruijnen T, Stemkens B, Terhaard CHJ, Lagendijk JJW, Raaijmakers CPJ, Tijssen RHN. Intrafraction motion quantification and planning target volume margin determination of head-and-neck tumors using cine magnetic resonance imaging. Radiother Oncol 2019;130:82–8. https://doi.org/10.1016/j.radonc.2018.09.015.

[34] Cheung JP, Park PC, Court LE, Ronald Zhu X, Kudchadker RJ, Frank SJ, et al. A novel dose-based positioning method for CT image-guided proton therapy. Med Phys 2013;40:1–9. https://doi.org/10.1118/1.4801910.

[35] Li B, Lee HC, Duan X, Shen C, Zhou L, Jia X, et al. Comprehensive analysis of proton range uncertainties related to stopping-power-ratio estimation using dual-energy CT imaging. Phys Med Biol 2017;62:7056–74. https://doi.org/10.1088/1361-6560/aa7dc9.

[36] Wagenaar D, Kierkels RGJ, Free J, Langendijk JA, Both S, Korevaar EW. Composite minimax robust optimization of VMAT improves target coverage and reduces non-target dose in head and neck cancer patients. Radiother Oncol 2019;136:71–7. https://doi.org/10.1016/j.radonc.2019.03.019.

TABLES AND FIGURES CAPTIONS

Table 1 - Study population characteristics (N = 10)

* The reported treatment modalities were classified as in the normal-tissue complication probability model for tube-feeding dependence.[15]

Table 2 - Average dose of all treatment scenarios including known error distributions for plans created with different robustness setup and range uncertainty criteria (N=10).

PC: Pharyngeal constrictor

#: For Δ/mm and Δ/% the slope of the regression fit is reported

*: $p<0.05$ in a two-tailed paired student t-test.

**: $p<\alpha$ after adjusting α using Bonferroni's correction for multiple testing.

Figure 1 – Generation of adjusted treatment plans with various uncertainty settings.

The following workflow is used to create treatment plans with smaller setup and range uncertainty settings from the clinical plan. The graphs show the dose profile near the targets for the dose distributions generated during the adjusted treatment plan generation. a) First, the clinical dose is calculated as the nominal scenario of the clinically used treatment plan which was robustly optimized with a setup uncertainty setting of 5 mm and range uncertainty setting of 3%. b) Next, a voxel-wise minimum robustness (multi-scenario) evaluation is performed with a shift equal to the difference in robustness setting between the new treatment plan and the clinical treatment plan. In this example, a treatment plan with a 2 mm setup uncertainty setting is created with an identical range uncertainty setting as for the clinically used treatment plan (3%). Therefore, the shift used for the voxel-wise minimum scenarios is 3 mm and 0% range. The resulting dose distribution has the same dose fall-off but is shifted more towards the targets and has a slight underdosage in the target. c) The dose inside the target is overridden to the clinical dose as determined above (step a) to avoid potential underdosage. d) Using voxel-based dose mimicking optimization, a deliverable treatment plan is created with a very similar nominal dose distribution to the predicted dose distribution.

Figure 2 - Treatment scenario calculation workflow.

To calculate a single treatment scenario the following workflow was followed. To incorporate changes due to changing anatomy, doses were recalculated on the weekly verification CTs. All seven verification CTs were shifted with the same systematic setup and range errors which were sampled from their determined distributions. For each CT, multiple daily fraction errors were sampled from [illegible]up error distribution and applied. Each treatment simulation therefore consists

of 35 daily perturbed CTs. This procedure was repeated 25 times for each treatment plan, sampling different systematic and random errors for each treatment scenario.

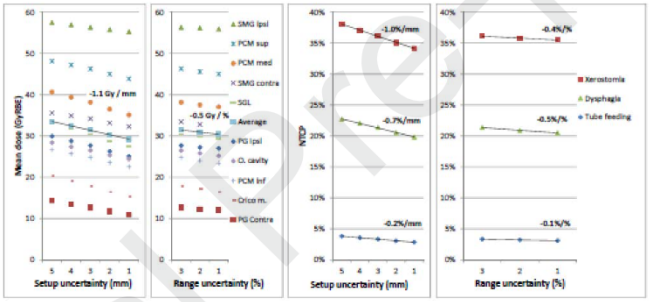
Figure 3 - Simulated clinical target volume dose and tumor control probability as a function of robustness setup and range uncertainty

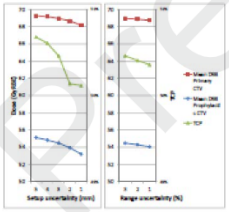
The reported doses incorporated known systematic and random setup and range uncertainties and anatomical changes.

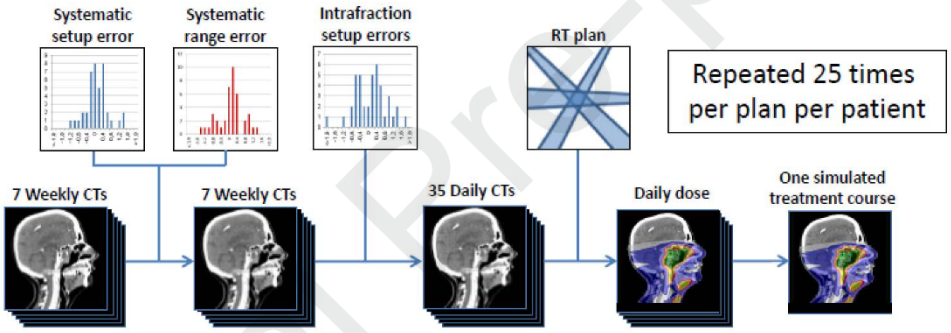
Figure 4 - Simulated organ at risk dose as a function of robustness setup and range uncertainty

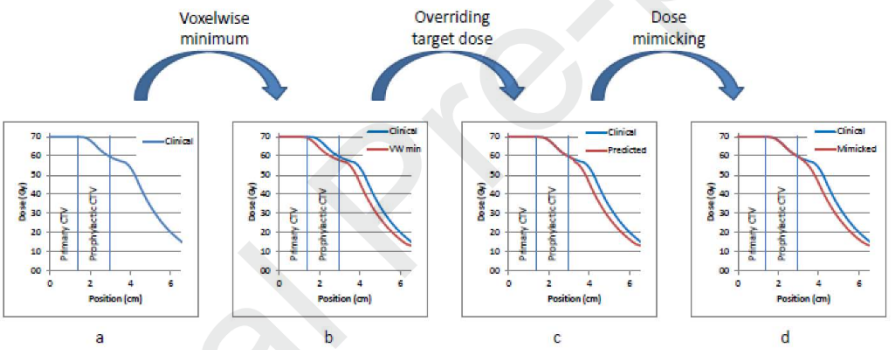
The reported doses incorporated known systematic and random setup and range uncertainties and anatomical changes.

SMG: Submandibular gland, PG: Parotid gland, ipsi: ipsilateral, contra: contralateral, PCM: Pharyngeal constrictor muscle, SGL: Supraglottic larynx, O. Cavity: Oral Cavity, Crico m.: Cricopharyngeal muscle


Mean dose (GyRBE)
-1.1 Gy / mm
Setup uncertainty (mm)
-0.5 Gy / %
Range uncertainty (%)
SMG ipsi
PCM sup
PCM med
SMG contra
SGL
Average
PG ipsi
Cl. cavity
PCM inf
Crico m.
PG Contre
NTCP
-1.0%/mm
-0.7%/mm
-0.2%/mm
Setup uncertainty (mm)
-0.4%/%
-0.5%/%
-0.1%/%
Range uncertainty (%)
Xerostomia
Dysphagia
Tube feeding




Systematic setup error
Systematic range error
Intrafraction setup errors
RT plan
Repeated 25 times per plan per patient
7 Weekly CTs
7 Weekly CTs
35 Daily CTs
Daily dose
One simulated treatment course


Voxelwise minimum
Overriding target dose
Dose mimicking
Clinical
VW min
Predicted
Mimicked
Primary CTV
Prophylactic CTV
Dose (Gy)
Position (cm)
a
b
c
d