

Aus der  
Klinik und Poliklinik für Strahlentherapie und Radioonkologie  
LMU Klinikum  
Ludwig-Maximilians-Universität München

Direktor: Prof. Dr. med. Claus Belka

**New approaches to improve treatment planning in  
radiotherapy with photons, protons and carbon ions**

Habilitationsschrift zur Erlangung der Venia Legendi im

Fach

Experimentelle Strahlentherapie

vorgelegt von

**Dr. rer. nat. Florian Kamp**

München 2024

## Habilitation Dr. rer. nat. Florian Kamp

This cumulative habilitation is primarily based on the following peer-reviewed scientific contributions (in the order of appearance in this thesis):

- [1] Hofmaier J, Dedes G, Carlson DJ, Parodi K, Belka C, **Kamp F**. Variance-based sensitivity analysis for uncertainties in proton therapy: A framework to assess the effect of simultaneous uncertainties in range, positioning and RBE model predictions on RBE-weighted dose distributions. *Med Phys*. 2020 Nov 19.
- [2] Hofmaier J, Walter F, Hadi I, Rottler M, von Bestenbostel R, Dedes G, Parodi K, Niyazi M, Belka C, **Kamp F**. Combining inter-observer variability, range and setup uncertainty in a variance-based sensitivity analysis for proton therapy. *Phys Imaging Radiat Oncol*. 2021 Dec 2;20:117-120.
- [3] **Kamp F**, Wilkens JJ. Application of variance-based uncertainty and sensitivity analysis to biological modeling in carbon ion treatment plans. *Med Phys*. 2019 Feb;46(2):437-447.
- [4] Hofmaier J, Haehnle J, Kurz C, Landry G, Maihoefer C, Schüttrumpf L, Süß P, Teichert K, Söhn M, Spahr N, Brachmann C, Weiler F, Thieke C, Küfer KH, Belka C, Parodi K, **Kamp F**. Multi-criterial patient positioning based on dose recalculation on scatter-corrected CBCT images. *Radiother Oncol*. 2017 Dec;125(3):464-469.
- [5] Neppl S, Kurz C, Köpl D, Yohannes I, Schneider M, Bondesson D, Rabe M, Belka C, Dietrich O, Landry G, Parodi K, **Kamp F**. Measurement-based range evaluation for quality assurance of CBCT-based dose calculations in adaptive proton therapy. *Med Phys*. 2021 Aug;48(8):4148-4159.
- [6] Neppl S, Landry G, Kurz C, Hansen DC, Hoyle B, Stöcklein S, Seidensticker M, Weller J, Belka C, Parodi K, **Kamp F**. Evaluation of proton and photon dose distributions recalculated on 2D and 3D U-net-generated pseudoCTs from T1-weighted MR head scans. *Acta Oncol*. 2019 Oct;58(10):1429-1434.
- [7] Rabe M, Thieke C, Düsberg M, Neppl S, Gerum S, Reiner M, Nicolay NH, Schlemmer HP, Debus J, Dinkel J, Landry G, Parodi K, Belka C, Kurz C\*, **Kamp F\***. Real-time 4DMRI-based internal target volume definition for moving lung tumors. *Med Phys*. 2020 Apr;47(4):1431-1442. \*Both authors contributed equally

Table of contents

1	Introduction and Background .....	3
2	Own scientific contributions.....	6
3	Conclusions and outlook .....	18
4	List of abbreviations .....	20
5	References.....	21
6	Acknowledgements .....	23
7	Declaration .....	24
8	Facsimile of relevant scientific contributions .....	25

## 1 Introduction and Background

Fractionated radiotherapy (RT), along with surgery, chemotherapy and hormonal or targeted therapy is one of the four modern classes of treatment of patients suffering from cancer. Today the majority of patients is treated with photons, and selected challenging cases with proton and carbon ion RT (the latter two are denoted as particle RT). The goal of RT is to treat the tumor with a therapeutic dose, while simultaneously sparing organs-at-risk (OARs) from radiation. Falling short of this objective may have severe consequences for the patient; thus treatment related uncertainties are a major concern in RT. Potential sources of uncertainty can be subdivided into two groups:

General RT (both photon and particle RT) (van Herk 2004, Hodapp 2012):

- inter-fractional anatomical changes of the patient (e.g. due to weight loss or tumor shrinkage)
- intra-fractional anatomical changes of the patient (moving of organs due to e.g. breathing)
- inter-fractional motion (uncertainty in the patient setup)
- delineation inaccuracies of targets as well as OARs (e.g. inter-observer variability (IOV))

Particle RT specific (Lomax 2008, Yang *et al* 2012, Paganetti 2014):

- range uncertainties (calibration curve for relative stopping power)
- predictions of spatially varying relative biological effectiveness (RBE) compared to photons

Keeping these uncertainties in mind, modern RT relies on margin-based target concepts (Hodapp 2012). The gross-tumor volume (GTV) is the gross visible extent of a tumor. Due to the high density of the cancer cells in the GTV, an adequate dose must be delivered to the whole GTV. It is extended by a margin to a clinical target volume (CTV), which may additionally contain microscopic, subclinical extensions of the tumor, which are not readily visible with medical imaging. In practice, to ensure that the CTV receives a dose that does not deviate significantly from the prescribed and planned doses, additional margins must be applied to the CTV accounting for internal changes of the patient, as well as for variations in patient position and beam geometries, both intra-fractionally and inter-fractionally. This leads to the concept of planning target volume (PTV). In clinical practice photon PTV margins are often isotropic extensions of the CTV. In contrast to this, due to the physical properties of charged particles, particle RT margin concepts are typically dependent on e.g. beam angle and geometry (Hodapp 2012, Schuemann *et al* 2014).

Substantial effort is invested to minimize the effect of the above listed uncertainties and hence to improve accuracy and reliability in RT. For example image-guided RT (IGRT) workflows using cone-beam computed tomography (CBCT) imaging at the treatment table are clinical standard for photon RT and are installed in an increasing number of particle RT centers worldwide. Currently these CBCT images are solely used to position the patient based on the visible anatomy. The CBCT images



themselves are not suitable for dose calculation and a (daily or online) treatment adaption on them is hence not feasible. In the scope of this habilitation an intensity correction for CBCT images, allowing for proton and photon dose calculation, was established and used to calculate the dose on the patient geometry of the day. The feasibility of a dose-guided patient positioning as an intermediate step to online adaptive RT (ART) was shown.

Recently, integrated magnetic resonance-linear accelerators (MR-Linacs, (Lagendijk *et al* 2014b)) became clinically available. The superior soft-tissue contrast allows for accurate visualization of targets and OARs at no imaging dose before and during the treatment. The vendors of the two certified MR-Linacs (ViewRay MRIdian (Mutic and Dempsey 2014) and Elekta Unity (Lagendijk *et al* 2014a)) provide online ART workflows (e.g. (Acharya *et al* 2016)) based on a pre-treatment MR images. Additionally it is possible to continuously image the patient during irradiation: 2D cine imaging allows tracking of the tumor with a very high temporal resolution. Even gated treatments are possible based on these cine MR images, meaning that the tumor is only irradiated while it is in a predefined location. This is especially of interest for intra-fractionally moving tumors located in e.g. lung or liver. In the scope of this habilitation, an approach to generate pseudo computed tomography images from MR images was evaluated. This is crucial since a direct dose calculation is not possible on MR images. The electron density of the different voxels has to be superimposed in a fast and reliable way in the online ART workflow. Additionally research was performed on repeated 4D-MR imaging, which was used to generate probability-of-presence (POP) maps for lung tumors. In a subsequent step the POP maps were compared to 4D CT based clinical target volume concepts.

In particle RT new developments such as range verification and improved computed tomography (CT) imaging for treatment planning (particle CT (Arbor *et al* 2015)), dual energy CT (van Elmpt *et al* 2016)) aim at minimizing residual range uncertainties. In terms of particle RT treatment planning, robust optimization (Unkelbach *et al* 2018) has brought a major improvement. Conceptually, robust optimization was so far only introduced for setup and range uncertainties, whereas uncertainties in RBE prediction and interplay between range, setup and RBE uncertainties have not been included yet. In order to fill this gap a variance-based statistical uncertainty and sensitivity approach (Saltelli 2007) was applied to particle RT in the scope of this habilitation. Global variance-based uncertainty and sensitivity analyses (abbreviated as UA and SA, respectively) can be used to assess the impact of uncertain input parameters on a model output. By assigning suitable probability distributions to the inputs and by a subsequent frequent recalculation of the model with inputs randomly and independently sampled from their corresponding distributions, a comprehensive UA and SA were achieved. Note that the model can be any analytical or numerical function (e.g. dose distributions, dose-volume histograms (DVHs) or any dose statistics) and that any probability density function can be employed in the random sampling to model the uncertainties. This includes the common normal, uniform or beta distributions but also discretized decisions between different scenarios. This statistical and computational flexibility was adapted to particle RT in order to offer

new treatment plan evaluation possibilities with the eventual goal to include UA and SA comprehensively in treatment plan optimization.

The above mentioned technical developments and improvements in accuracy eventually aim at reducing needed margin size. This reduction is the most efficient way to spare OARs, since a smaller tissue volume has to be irradiated at prescribed dose levels. To summarize, the research performed in the scope of this habilitation intends to pave the way towards new approaches to improve treatment planning in photon and particle RT by quantifying, accounting and minimizing the different uncertainties. To this aim, the following aspects have been addressed

1. Sensitivity and uncertainty analysis for particle RT treatment planning (section 2.1)
2. CBCT intensity correction to facilitate dose calculation on the patient geometry of the treatment day (section 2.2)
3. MR-based RT for brain tumor and moving lung tumors (section 2.3)

In the following chapters the selected peer-reviewed publications as first or last author in the scope of this habilitation are presented in more detail. Facsimiles of these publications can be found in chapter 8.

## 2 Own scientific contributions

### 2.1 Projects on sensitivity and uncertainty analysis in particle therapy

In order to account for, quantify and evaluate the above mentioned uncertainties in particle therapy a statistical calculation framework was developed and implemented in the last years. The core is a Monte Carlo derived, variance-based uncertainty and sensitivity analysis (UA and SA) to quantify the impact of different biological and physical uncertainties on particle therapy treatment plans. The UA and SA framework was added to the particle RT extension (Schell and Wilkens 2010) of the research treatment planning system CERR (a Computational Environment for Radiotherapy Research) (Deasy *et al* 2003).

The underlying principle is the repeated evaluation of a function (here an RBE-weighted dose (RWD) distribution or a RBE-weighted DVH (RW-DVH) of a treatment plan) for up to  $10^5$  times. For each of those runs all potentially uncertain inputs (e.g. the dose, isocenter position and biological modeling parameters) are simultaneously sampled randomly from their assigned distributions. This facilitates a ranking of the input parameter/uncertainty pairs according to their impact on the uncertainty of the result using variance-based statistics.

In the last years this SA approach has been successfully applied to proton and carbon ion RT.

#### 2.1.1 Hofmaier J, Dedes G, Carlson DJ, Parodi K, Belka C, **Kamp F**. Variance-based sensitivity analysis for uncertainties in proton therapy: A framework to assess the effect of simultaneous uncertainties in range, positioning and RBE model predictions on RBE-weighted dose distributions. *Med Phys*. 2020 Nov 19

The variance-based SA requires a large number ( $10^4$ - $10^5$ ) of RWD calculations. Based on a particle RT extension of the research treatment planning system CERR a fast, graphics processing unit (GPU) accelerated pencil beam modeling of patient and range shifts was implemented. In addition to the setup and range uncertainties proton therapy is affected by uncertainties in RBE prediction. While to date a constant RBE of 1.1 is commonly assumed clinically, the actual RBE is known to increase towards the distal end of the spread-out Bragg peak. Several models for variable RBE predictions exist. In this study two biological models were included: The mechanistic repair-misrepair-fixation (RMF) model (Carlson *et al* 2008), A2) and the phenomenological Wedenberg model (Wedenberg *et al* 2013). Here, the input parameters (patient position, proton range, RBE model parameters) were sampled simultaneously from their assumed probability distributions. Statistical formalisms rank the input parameters according to their influence on the overall uncertainty of RW-DVH quantiles and the RWD in every voxel, resulting in relative normalized sensitivity indices (“S=0: non influential input”; “S=1: the only influential input”). Results were visualized as RW-DVHs with uncertainty bars and sensitivity maps (figure 1).

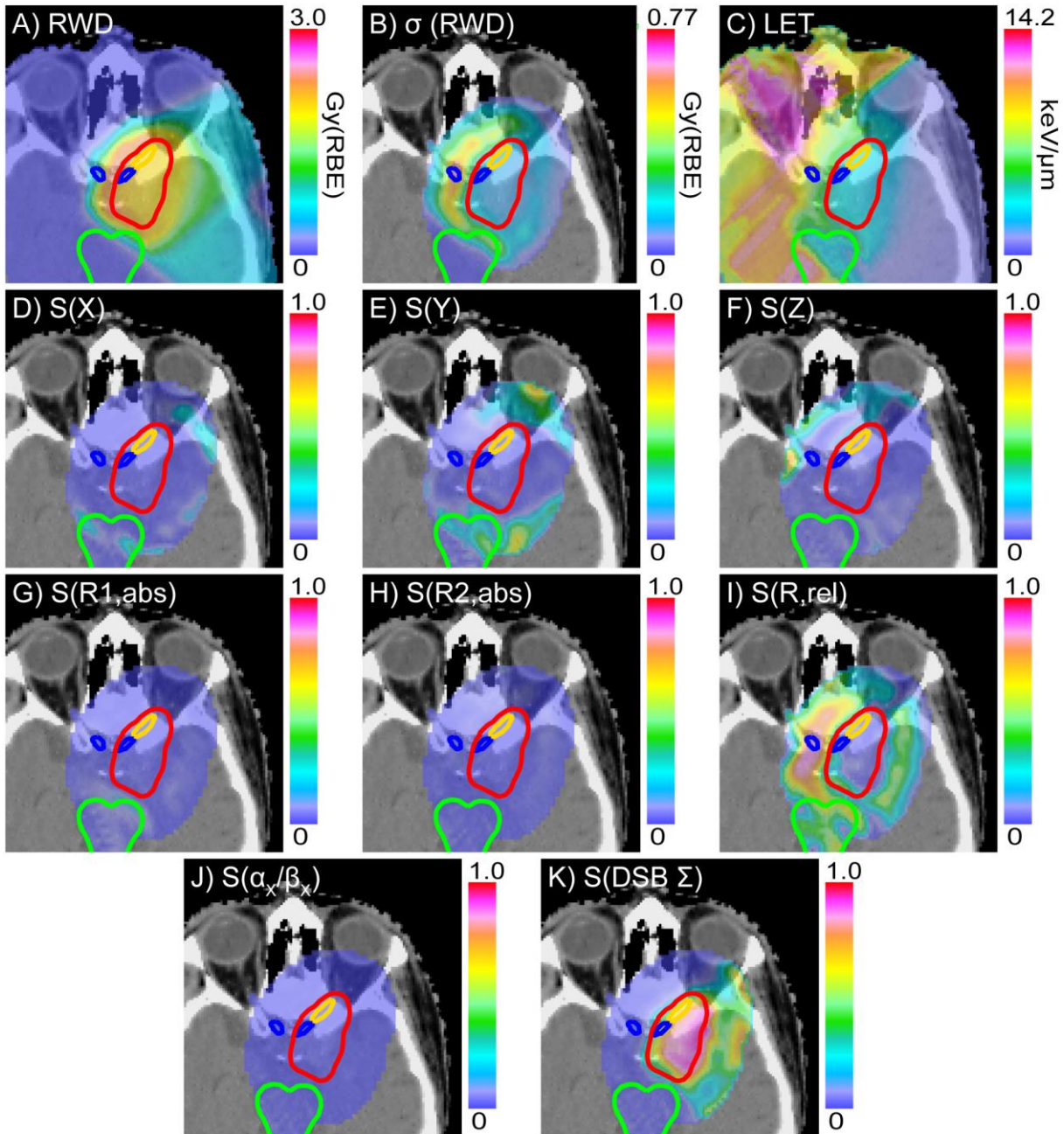


Figure 1 : A) RWD distribution, B) local standard deviation, C) LET distribution and sensitivity maps (D-K) for an exemplary patient and the RMF model in the calculation with  $\alpha_x / \beta_x = 2$  Gy with a standard deviation of 10 %. Figure adapted from [1].

The approach was demonstrated for two representative brain tumor cases and a prostate case. The full SA including  $\sim 2.8 \cdot 10^4$  RWD calculations took 39, 11 and 55 minutes, respectively. Range uncertainty was an important contributor to the overall uncertainty at the distal end of the target region. The smaller uncertainty inside the target itself was governed by the impact of biological uncertainties. Consequently, the uncertainty of the RW-DVH quantile  $D_{98\%}$  for target was dominated by range uncertainty while the uncertainty of the mean target dose was caused predominantly by the uncertainty in biological parameters. It was shown that the SA framework is a powerful and flexible tool to evaluate uncertainty in RWD distributions and DVH quantiles, taking into account physical and RBE uncertainties and their interplay. The

additionally gained information will have implications for future approaches to biologically robust planning and optimization.

2.1.2 Hofmaier J, Walter F, Hadi I, Rottler M, von Bestenbostel R, Dedes G, Parodi K, Niyazi M, Belka C, **Kamp F**. Combining inter-observer variability, range and setup uncertainty in a variance-based sensitivity analysis for proton therapy. *Phys Imaging Radiat Oncol*. 2021 Dec 2;20:117-120.

In this project the UA and SA of the previous chapter was extended to uncertain CTVs. Due to IOV (Hellebust *et al* 2013, Lobefalo *et al* 2013, Eminowicz *et al* 2016, Vinod *et al* 2016) the target structure in RT is inherently uncertain. The feasibility to extend the UA and SA by sampling over discrete scenarios (different CTVs) in one UA and SA was shown successfully. The presented framework is based on the particle extension (Schell and Wilkens 2010) of CERR (Deasy *et al* 2003) combined with the fast GPU-based pencil beam RWD calculation algorithm. In order to demonstrate the combined impact of IOV, setup and range uncertainty in a variance-based SA ten patients with skull base meningioma were evaluated. For each patient, four clinicians independently delineated the GTV taking into account all imaging modalities. A consensus GTV was created using the simultaneous truth and performance level estimation (STAPLE) algorithm (Warfield *et al* 2004). Following current guidelines no margin was added to create the CTVs (i.e. CTV = GTV). The mean calculation time to perform the SA including  $1.6 \cdot 10^4$  RWD recalculations was 59 min. For two patients in this dataset, IOV had a relevant impact on the estimated CTV  $D_{95\%}$  uncertainty.

In this project an important additional source of uncertainty was included in the analysis and evaluated in the context of several uncertain parameters. From a more technical point of view the feasibility to add the sampling over discrete IOV scenarios to the SA was shown. This is an important step to comprehensively simulate and evaluate all major sources of uncertainty in (particle therapy) treatment planning and plan evaluation.

2.1.3 **Kamp F**, Wilkens JJ. Application of variance-based uncertainty and sensitivity analysis to biological modeling in carbon ion treatment plans. *Med Phys*. 2019 Feb;46(2):437-447.

In carbon ion therapy the RBE itself as well as the impact of biological uncertainties on RBE predictions is higher than in proton therapy. The variance-based UA and SA was implemented to quantify and hence access the impact of these biological uncertainties on RWD distributions in carbon ion therapy treatment planning.

Based on an exemplary astrocytoma patient case, the application of variance-based SA for biological measures was demonstrated and applied to a two-field spot scanning carbon ion treatment plan for two commonly used biological models (Scholz *et al* 1997, Carlson *et al* 2008), A2, A3) and two representative tissue parameter sets. A voxel-wise calculation for  $2.9 \cdot 10^5$  voxels took ~6 h. A structure-based SA, which adds an uncertainty band to a RW-DVH and shows how to decrease the overall



uncertainty in the most effective way, can be calculated in 0.1–1.5 h (depending on the size of the structure). The uncertainties in RBE, RWD or RW-DVH were broken down to the contributions of different uncertainties in the (biological) model input. Biological uncertainties have a higher impact on the resulting RBE and RWD than uncertainties in the physical dose. Excluding the physical dose from the SA only slightly decreased the overall uncertainty, emphasizing the necessity to comprehensively include biological uncertainties into treatment plan evaluation.

The way to this study was paved by two preliminary studies (A2, A3). Within the first study, RMF model predictions were combined with fragmentation spectra to perform carbon ion treatment planning including RBE model predictions (A2). The second publication was a technical note that demonstrated the decoupling of physical beam properties and biological input in the frame of the RMF-model, facilitating very fast RBE model predictions and, even more important, very fast changes of biological tissue properties as input to the RBE prediction (A3).

The method was presented and discussed based on an exemplary patient case. Future applications include sensitivity-based, biologically robust, treatment plan optimization using the newly accessible, patient-specific uncertainty and sensitivity information.

#### 2.1.4 Further publications related to sensitivity and uncertainty analysis and modeling of the relative biological effectiveness

- A1. Zvereva A, **Kamp F**, Schlattl H, Zankl M, Parodi K. Impact of interpatient variability on organ dose estimates according to MIRDO schema: Uncertainty and variance-based sensitivity analysis. *Med Phys*. 2018 Jul;45(7):3391-3403.
- A2. **Kamp F**, Cabal G, Mairani A, Parodi K, Wilkens JJ, Carlson DJ. Fast Biological Modeling for Voxel-based Heavy Ion Treatment Planning Using the Mechanistic Repair-Misrepair-Fixation Model and Nuclear Fragment Spectra. *Int J Radiat Oncol Biol Phys*. 2015 Nov 1;93(3):557-68.
- A3. **Kamp F**, Carlson DJ, Wilkens JJ. Rapid implementation of the repair-misrepair-fixation (RMF) model facilitating online adaption of radiosensitivity parameters in ion therapy. *Phys Med Biol*. 2017 Jul 7;62(13):N285-N296.
- A4. **Kamp F**, Brüningk S, Cabal G, Mairani A, Parodi K, Wilkens JJ. Variance-based sensitivity analysis of biological uncertainties in carbon ion therapy. *Phys Med*. 2014 Jul;30(5):583-7.
- A5. Meschini G, **Kamp F**, Hofmaier J, Reiner M, Sharp G, Paganetti H, Belka C, Wilkens JJ, Carlson DJ, Parodi K, Baroni G, Riboldi M. Modeling RBE-weighted dose variations in irregularly moving abdominal targets treated with carbon ion beams. *Med Phys*. 2020 Jul;47(7):2768-2778.
- A6. Resch AF, Landry G, **Kamp F**, Cabal G, Belka C, Wilkens JJ, Parodi K, Dedes G. Quantification of the uncertainties of a biological model and their impact on

variable RBE proton treatment plan optimization. *Phys Med.* 2017 Apr;36:91-102.

## 2.2 CBCT based adaptive radiotherapy for photon and proton radiotherapy treatment

In the last years a method for CBCT intensity correction was established at the LMU in Munich. In close cooperation with the Medical Physics chair of the LMU physics faculty, a method to correct CBCT intensity was implemented and evaluated. This allows reliable and accurate photon as well as proton dose calculation on the patient geometry of the day based on the CBCT.

The CBCT images are corrected for typical CBCT artefacts and converted to planning CT (pCT) equivalent Hounsfield unit (HU) numbers by using the method proposed by (Park *et al* 2015). Briefly, the method makes use of a virtual CT (vCT) obtained by deformable image registration (DIR) of the pCT to a CBCT image. The vCT is used to generate idealized projections which allow estimation and correction of non-idealities in the original CBCT projections. This approach has been extensively evaluated in terms of HU as well as photon and proton RT dose calculations (Park *et al* 2015)([4],B1,B2,B3) and used as input for several publications about artificial intelligence approaches for CBCT intensity correction (B4,B5,B7). The process eventually yields an intensity corrected CBCT image,  $CBCT_{cor}$ , reconstructed from the corrected projections. An example of the method is shown in figure 2.  $CBCT_{cor}$  displays the patient geometry of the day (as does the CBCT) but with CT values comparable to the pCT, allowing reliable dose calculations for both proton and photon RT. The generation of one  $CBCT_{cor}$  takes about 5-10 min with few manual steps in the process. The usage of a DIR and resulting vector field has the advantage that contoured structures of the pCT can be warped to the daily CBCT geometry. This provides an automated contour suggestion as starting point for the review by a physician.

### 2.2.1 Hofmaier J, Haehnle J, Kurz C, Landry G, Maihoefer C, Schüttrumpf L, Süss P, Teichert K, Söhn M, Spahr N, Brachmann C, Weiler F, Thieke C, Küfer KH, Belka C, Parodi K, **Kamp F**. Multi-criterial patient positioning based on dose recalculation on scatter-corrected CBCT images. *Radiother Oncol.* 2017 Dec;125(3):464-469.

In this project the  $CBCT_{cor}$  was used to evaluate the feasibility and potential advantages of dose-guided patient positioning based on dose recalculation on the patient geometry of the day based on CBCT. The dose-guided patient positioning is a multi-criteria approach previously introduced by Haehnle et al. (B9). In the conducted project the scatter correction approach has been employed to facilitate precise photon dose calculations on daily CBCT images. The proposed tool for interactive multi-criterial dose-guided patient positioning which uses interpolation between pre-calculated sample doses has been utilized. The workflow was retrospectively evaluated for two head and neck patients with a total of 39 CBCTs.

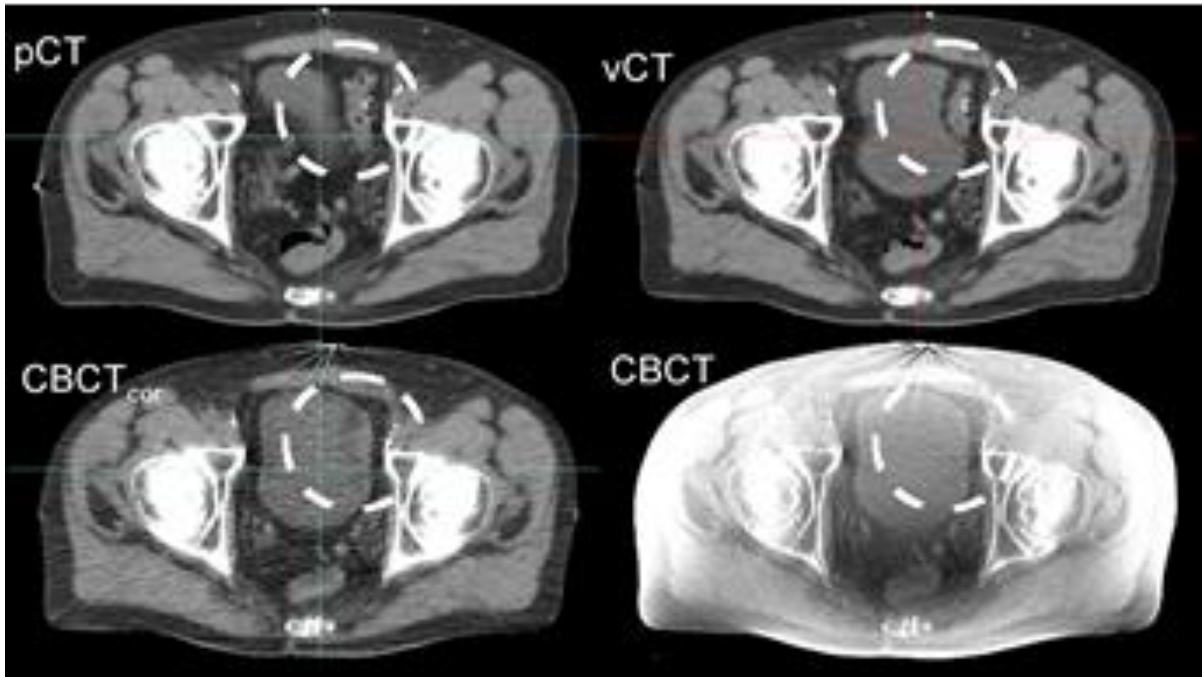


Figure 2: The vCT approach fails to properly deform the pCT to match the CBCT, as seen in the region indicated by the white circle for a prostate patient case. The scatter corrected CBCT (CBCT<sub>cor</sub>) shows similar image quality as the pCT and the correct anatomy when compared to the CBCT. The same window and level has been used for displaying. Figure adapted from (B1).

DVH parameters were compared to rigid image registration based isocenter corrections, mimicking the clinical scenario. The accuracy of the dose interpolation was successfully compared against Monte Carlo dose calculation facilitating the implementation of dose guided patient positioning. Compared to the clinical scenario, the mean dose to the parotid glands could be improved for 2 out of 5 fractions for the first patient while other parameters were preserved. For the second patient, the mean coverage over all fractions of the high dose PTV could be improved by 4%. For this patient, dose coverage improvements had to be traded against OAR doses within their clinical tolerance limits. In conclusion, it was shown that dose guided patient positioning for photon RT is feasible and offers increased control over target dose coverage and doses to OARs. The presented dose-guided patient positioning approach has advantages over online ART workflows, since the treatment plan is not re-optimized and hence has not to be approved again by a clinician and a medical physicist. Note that the concept of dose-guided patient positioning has also been tested for proton RT (B10).

2.2.2 Neppl S, Kurz C, Köpl D, Yohannes I, Schneider M, Bondesson D, Rabe M, Belka C, Dietrich O, Landry G, Parodi K, **Kamp F**. Measurement-based range evaluation for quality assurance of CBCT-based dose calculations in adaptive proton therapy. *Med Phys*. 2021 Aug;48(8):4148-4159

Adaptive proton RT based on volumetric in-room imaging requires commissioning and quality assurance measures for the employed imaging. This is especially crucial in case of online adaptive scenarios. In this publication a quality assurance method for 3D range verification in CBCT<sub>cor</sub> based dose calculations was introduced using dosimetry gel and films in 3D printed head phantoms. The aim was to explicitly



account for the impact of the treatment plan re-optimization process. Patient specific 3D printed head phantoms based on the clinical pCT were available for three patient cases. For these patients single field uniform dose pencil beam scanning proton plans were created on both pCT and CBCT<sub>cor</sub>, respectively. Here the same target volumes and optimization constraints were used for pCT and CBCT<sub>cor</sub>. The resulting treatment plans and their respective forward recalculations on pCT or CBCT<sub>cor</sub> were compared in terms of proton range differences (80% distal fall-off). The two treatment plans per patient were additionally irradiated and the resulting dose distribution measured using film and gel dosimetry inside the 3D printed head phantoms. In order to extract the impact of the imaging, the simulated as well as the measured range differences were corrected for range differences originating from the initial plans.

The results indicate that a CBCT<sub>cor</sub> based online adaptive proton therapy workflow for head irradiations is feasible. The here introduced measurement- and simulation-based method was equivalent to the standard recalculation approach commonly found in the literature. The advantage is that it additionally has the capability to catch effects of image differences on the optimization step of the workflow. The observed uncertainties could be kept within those of the image registration and positioning. The proposed validation approach is transferable to other potential in-room imaging modalities, e.g. to MR-based pseudoCTs.

### 2.2.3 Further publications related to CBCT based adaptive radiotherapy for photon and proton radiotherapy treatment and dose-guided patient positioning

- B1. Kurz C, **Kamp F**, Park YK, Zöllner C, Rit S, Hansen D, Podesta M, Sharp GC, Li M, Reiner M, Hofmaier J, Nepl S, Thieke C, Nijhuis R, Ganswindt U, Belka C, Winey BA, Parodi K, Landry G. Investigating deformable image registration and scatter correction for CBCT-based dose calculation in adaptive IMPT. *Med Phys*. 2016 Oct;43(10):5635.
- B2. Schmitz H, Rabe M, Janssens G, Bondesson D, Rit S, Parodi K, Belka C, Dinkel J, Kurz C, **Kamp F\***, Landry G\*. Validation of proton dose calculation on scatter corrected 4D cone beam computed tomography using a porcine lung phantom. *Phys Med Biol*. 2021 Aug 30;66(17). \*Both authors contributed equally
- B3. Bondesson D, Meijers A, Janssens G, Rit S, Rabe M, **Kamp F**, Niepel K, Otter LAD, Both S, Brousmiche S, Dinkel J, Belka C, Parodi K, Knopf A, Kurz C, Landry G. Anthropomorphic lung phantom based validation of in-room proton therapy 4D-CBCT image correction for dose calculation. *Z Med Phys*. 2020 Nov 25:S0939-3889(20)30099-4
- B4. Kurz C, Maspero M, Savenije MHF, Landry G, **Kamp F**, Pinto M, Li M, Parodi K, Belka C, Van den Berg CAT. CBCT correction using a cycle-consistent generative adversarial network and unpaired training to enable photon and proton dose calculation. *Phys Med Biol*. 2019 Nov 15;64(22):225004

- B5. Landry G, Hansen D, **Kamp F**, Li M, Hoyle B, Weller J, Parodi K, Belka C, Kurz C. Comparing Unet training with three different datasets to correct CBCT images for prostate radiotherapy dose calculations. *Phys Med Biol*. 2019 Jan 24;64(3):035011.
- B6. Niepel K, **Kamp F**, Kurz C, Hansen D, Rit S, Nepl S, Hofmaier J, Bondesson D, Thieke C, Dinkel J, Belka C, Parodi K, Landry G. Feasibility of 4DCBCT-based proton dose calculation: An ex vivo porcine lung phantom study. *Z Med Phys*. 2019 Aug;29(3):249-261.
- B7. Hansen DC, Landry G, **Kamp F**, Li M, Belka C, Parodi K, Kurz C. ScatterNet: A convolutional neural network for cone-beam CT intensity correction. *Med Phys*. 2018 Nov;45(11):4916-4926.
- B8. Zöllner C, Rit S, Kurz C, Vilches-Freixas G, **Kamp F**, Dedes G, Belka C, Parodi K, Landry G, Decomposing a prior-CT-based cone-beam CT projection correction algorithm into scatter and beam hardening components, *Phys. Imag. Radiat. Oncol. (phiRO)* 3, 49-52.
- B9. Haehnle J, Süß P, Landry G, Teichert K, Hille L, Hofmaier J, Nowak D, **Kamp F**, Reiner M, Thieke C, Ganswindt U, Belka C, Parodi K, Küfer KH, Kurz C. A novel method for interactive multi-objective dose-guided patient positioning. *Phys Med Biol*. 2017 Jan 7;62(1):165-185
- B10. Kurz C, Süß P, Arnsmeier C, Haehnle J, Teichert K, Landry G, Hofmaier J, Exner F, Hille L, **Kamp F**, Thieke C, Ganswindt U, Valentini C, Hölscher T, Troost E, Krause M, Belka C, Küfer KH, Parodi K, Richter C. Dose-guided patient positioning in proton radiotherapy using multicriteria-optimization. *Z Med Phys*. 2019 Aug;29(3):216-228.

## 2.3 Application of magnetic resonance imaging in treatment planning

- 2.3.1 Nepl S, Landry G, Kurz C, Hansen DC, Hoyle B, Stöcklein S, Seidensticker M, Weller J, Belka C, Parodi K, **Kamp F**. Evaluation of proton and photon dose distributions recalculated on 2D and 3D Unet-generated pseudoCTs from T1-weighted MR head scans. *Acta Oncol*. 2019 Oct;58(10):1429-1434.

In a subsequent step the ScatterNet employed for CBCT intensity correction (e.g. B7) of the previous chapter has been extended to generate pseudoCTs from cranial MR image. The underlying clinical scenario is either a so-called MR-only workflow or an MR-guided RT workflow. In order to perform dose calculations, in both workflows it is necessary to convert MR images to pseudoCT images, since the electron density information is required for dose calculations.

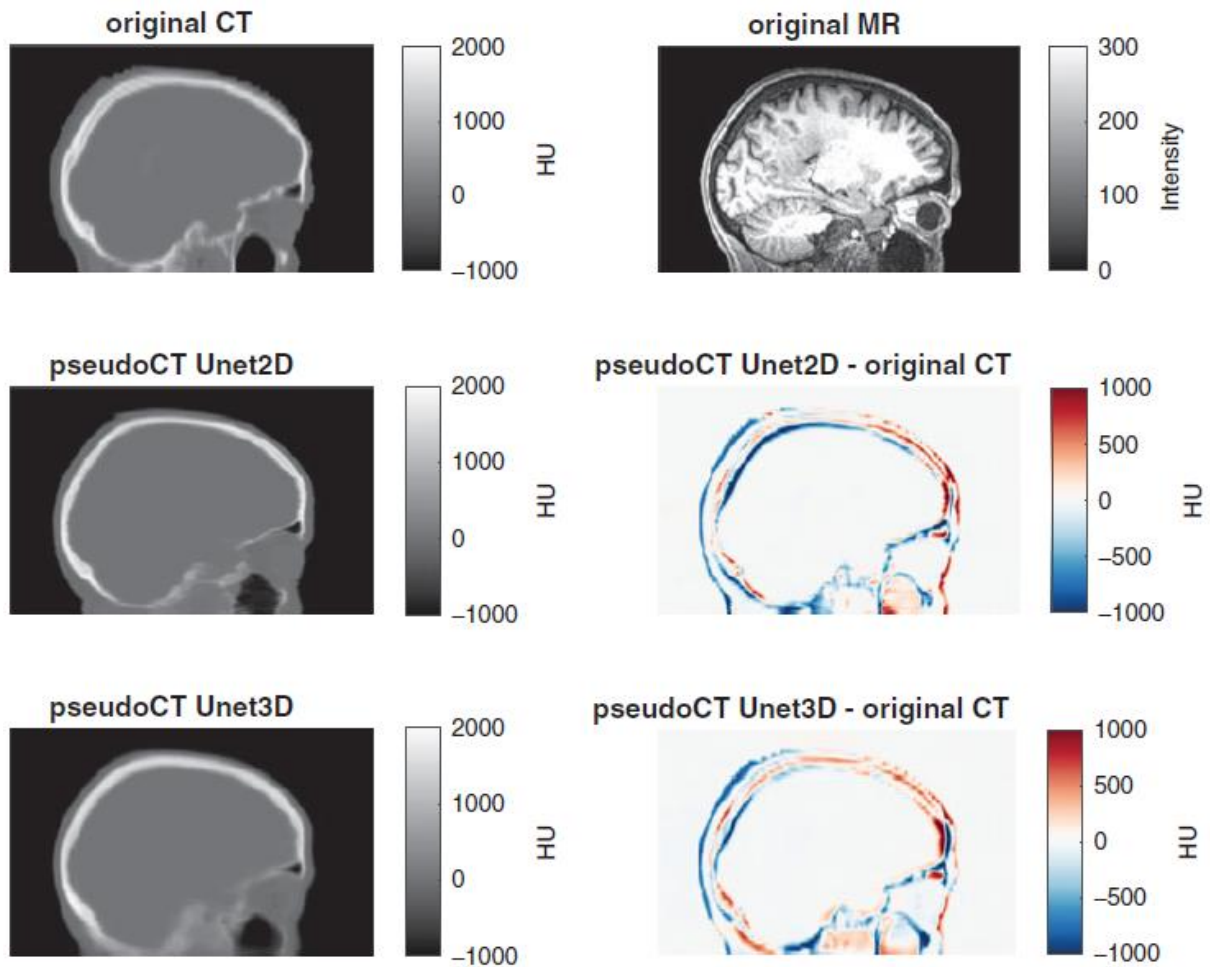


Figure 3: Sagittal views for an exemplary patient of the application set of the CT (upper left) and the MR (upper right). The corresponding slices of the 2D trained pseudoCT Unet2D (middle left), and the 3D trained pseudoCT Unet3D (lower left) are displayed. The difference plots “pseudoCT Unet2D – original CT” (middle right) and “pseudoCT Unet3D – original CT” (lower right) are also shown. Figure adapted from [6]. It can be seen that slice discontinuities are reduced in the 3D case at the cost of sharpness.

Recently several groups performed this pseudoCT generation using artificial intelligence (Han 2017, Dinkla *et al* 2018, Jin *et al* 2019). These studies as well as the project of this habilitation were motivated by a need of fast and reliable conversion of MR images to X-ray CT images, especially in the scope of recent developments of MR based adaptive strategies for photon and, potentially for proton RT. As stated above, precise CT values are needed for photon and proton dose calculation. A major step compared to the ScatterNet introduced by Hansen *et al.* (B7) was the achieved extension to 3D to reduce inter-slice discontinuities seen in its 2D application. Consequently two U-shaped convolutional neural networks (Unet) were implemented. The results of the so-called Unet2D as well as the Unet3D pseudoCT generation were compared against each other. A database of 89 T1-weighted MR head scans with about 100 slices each, including rigidly registered CTs, was created. Twenty-eight validation patient datasets were randomly sampled, and four patient datasets were selected for application. The remaining patient datasets were used to train both Unet2D as well as Unet3D. A stack size of 32 slices was used for 3D training. For all application dataset cases, volumetric modulated arc RT photon and single-field uniform dose pencil-beam scanning proton plans at four different gantry

angles were optimized for a generic target on the CT and recalculated on 2D and 3D Unet-based pseudoCTs. Mean (absolute) error (MAE/ME) and a gradient sharpness estimate were used to quantify the image quality. Three-dimensional gamma and dose difference analyses were performed for photon (gamma criteria: 1%, 1mm) and proton dose distributions (gamma criteria: 2%, 2mm). Range (80% fall off) differences for beam's eye view profiles were evaluated for protons. Training 36 h for 1000 epochs in 3D or 6 h for 200 epochs in 2D yielded a maximum MAE of 147 HU and 135 HU, respectively, for the application patient datasets. An exemplary result adopted from this publication [6] is shown in figure 3. Except for one patient gamma pass rates for photon and proton dose distributions were above 96% for both Unets. Slice discontinuities were reduced for 3D training at the cost of sharpness. Image analysis revealed a slight advantage of 2D Unets compared to 3D Unets. Similar dose calculation performance was reached for the 2D and 3D network. Overall, good accuracy was found for both proton and for photon dose calculation and a considered cohort of brain tumor patients.

2.3.2 Rabe M, Thieke C, Düsberg M, Nepl S, Gerum S, Reiner M, Nicolay NH, Schlemmer HP, Debus J, Dinkel J, Landry G, Parodi K, Belka C, Kurz C\*, **Kamp F\***. Real-time 4DMRI-based internal target volume definition for moving lung tumors. *Med Phys.* 2020 Apr;47(4):1431-1442. \*Both authors contributed equally

A further promising application of MR imaging in RT is the possibility to show moving patient geometries in 4D. Since MR imaging does not deposit dose in the patient (in contrast to e.g. CBCT) it can be repeatedly used to image e.g. the respiratory induced tumor movement over time. In clinical practice respiratory-induced target motion can be accounted for by internal target volumes (ITV) (Hodapp 2012) or mid-ventilation target volumes (midV) (Wolthaus *et al* 2006) defined on the basis of a 4D-CT. Intrinsic limitations of these approaches can result in target volumes that are not representative for the GTV motion over the course of treatment. To address these limitations, a novel patient-specific ITV definition method based on real-time 4D magnetic resonance imaging (rt-4DMRI) was proposed. The method was evaluated based on three lung cancer patients who underwent weekly rt-4DMRI scans. A total of 24 datasets were included in this retrospective study. The GTV was contoured on breath-hold MR images and propagated to all rt-4DMRI images by deformable image registration. Different targets were created for the first (reference) imaging sessions: ITVs encompassing all GTV positions over the complete ( $ITV^{80s}$ ) or partial acquisition time ( $ITV^{10s}$ ), ITVs including only voxels with a GTV probability-of-presence (POP) of at least 5% ( $ITV^{5\%}$ ) or 10% ( $ITV^{10\%}$ ), and the mid-ventilation GTV position. An exemplary POP distribution on top of an rt-4DMRI is shown in figure 4. Reference planning target volumes ( $PTV_r$ ) were created by adding margins around the ITVs and midV target volumes. The geometrical overlap of the  $PTV_r$  with  $ITV_n^{5\%}$  from the six to eight subsequent imaging sessions on days  $n$  was quantified in terms of the Dice similarity coefficient (DSC), sensitivity (SE:  $(PTV_r / ITV_n^{5\%}) / ITV_n^{5\%}$ ) and precision (PRE:  $(PTV_r / ITV_n^{5\%}) / PTV_r$ ) as surrogates for target coverage and normal tissue sparing.

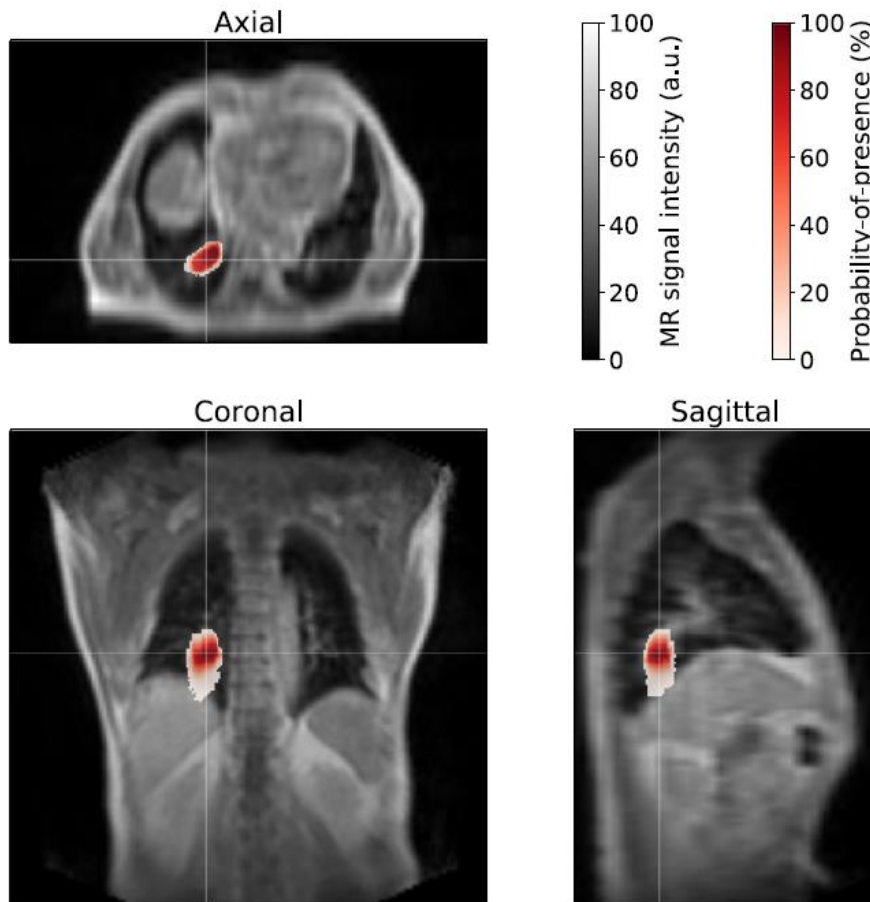


Figure 4: Three-dimensional probability-of-presence (POP) distribution displayed on top of the MR images of an exemplary patient. The POP is color-encoded and superimposed on four-dimensional-magnetic resonance imaging (MRI). The crosshairs indicate the slice positions in the respective axial, coronal, and sagittal views. Figure adapted from [7].

The results can be summarized as follows: patient-specific analysis yielded a high variance of the overlap values of  $PTV_r^{10s}$ , when different periods within the reference imaging session were sampled. The mid-ventilation-based PTVs were smaller than the ITV-based PTVs. Whereas the SE was high for patients with small breathing pattern variations, changes of the median breathing amplitudes in different imaging sessions led to inferior SE values for the mid-ventilation PTV for one patient. In contrast,  $PTV_r^{5\%}$  and  $PTV_r^{10\%}$  showed higher SE values with a higher robustness against inter-fractional changes, at the cost of larger target volumes. In conclusion, the results indicate that rt-4DMRI could be valuable for the definition of target volumes based on the GTV POP to achieve a higher robustness against inter-fractional changes than feasible with today's 4D-CT-based target definition concepts.

### 2.3.3 Further publications related to MR in radiotherapy

- C1. Rabe M, Paganelli C, Riboldi M, Bondesson D, Schneider MJ, Chmielewski T, Baroni G, Dinkel J, Reiner M, Landry G, Parodi K, Belka C, **Kamp F\***, Kurz C\*. Porcine lung phantom-based validation of estimated 4D-MRI using orthogonal

## Habilitation Dr. rer. nat. Florian Kamp

cine imaging for low-field MR-Linacs. *Phys Med Biol.* 2020 Nov 10. \*Both authors contributed equally

- C2. Kurz C, Buizza G, Landry G, **Kamp F**, Rabe M, Paganelli C, Baroni G, Reiner M, Keall PJ, van den Berg CAT, Riboldi M. Medical physics challenges in clinical MR-guided radiotherapy. *Radiat Oncol.* 2020 May 5;15(1):93.
- C3. Nierer L, Kamp F, Reiner M, Corradini S, Rabe M, Dietrich O, Parodi K, Belka C, Kurz C, Landry G. Evaluation of an anthropomorphic ion chamber and 3D gel dosimetry head phantom at a 0.35 T MR-linac using separate 1.5 T MR-scanners for gel readout. *Z Med Phys.* 2022 Mar 16:S0939-3889(22)00006-X.
- C4. Da Silva Mendes V, Nierer L, Li M, Corradini S, Reiner M, **Kamp F**, Niyazi M, Kurz C, Landry G, Belka C. Dosimetric comparison of MR-linac-based IMRT and conventional VMAT treatment plans for prostate cancer. *Radiat Oncol.* 2021 Jul 21;16(1):133.
- C5. Kroll C, Dietrich O, Bortfeldt J, **Kamp F**, Nepl S, Belka C, Parodi K, Baroni G, Paganelli C, Riboldi M. Integration of Spatial Distortion Effects in a 4D Computational Phantom for Simulation Studies in Extra-Cranial MRI-guided Radiation Therapy: Initial Results. *Med Phys.* 2020 Nov 21.
- C6. Kroll C, Dietrich O, Bortfeldt J, Paganelli C, Baroni G, **Kamp F**, Nepl S, Belka C, Parodi K, Opel M, Riboldi M. Improving the modelling of susceptibility-induced spatial distortions in MRI-guided extra-cranial radiotherapy. *Phys Med Biol.* 2019 Oct 10;64(20):205006.
- C7. Dumlu HS, Meschini G, Kurz C, **Kamp F**, Baroni G, Belka C, Paganelli C, Riboldi M. Dosimetric impact of geometric distortions in an MRI-only proton therapy workflow for lung, liver and pancreas. *Z Med Phys.* 2020 Nov 6:S0939-3889(20)30104-5.



### 3 Conclusions and outlook

In the scope of this habilitation, studies on several aspects of treatment planning and treatment plan evaluation for radiotherapy with photons and particles have been conducted. These, on the one hand, focused on methods how to model, propagate and interpret uncertainties in the treatment planning process and, on the other hand, aimed at providing new insight to methods and techniques useful for ART.

In the scope of this habilitation it was possible to show the power of the very flexible variance-based UA and SA approach. Future research will extend the present work in several directions. First, the method can be applied to photon RT, comprehensively modeling the impact of setup uncertainties, IOV and potentially also inter-fractional changes due to varying geometry of a patient during the course of a fractionated radiotherapy treatment. Here the intensity corrected CBCT images of the second part of the habilitation could be used. Second, the newly derived information on UA and SA could be included into the treatment plan optimization itself to achieve additional robustness in particle treatment plans, allowing accounting for uncertainties in setup, range and RBE prediction as well as IOV in the target definition. In order to achieve this, an optimization cost function including the UA and SA measures could be implemented in a research treatment planning platform. The framework can be used to evaluate and quantify the impact of improvements in terms of e.g. smaller range uncertainties coming with new technologies, e.g. dual energy CT or particle CT.

The discussion and outlook in terms of methods and steps towards ART has to be subdivided between photon and particle therapy. In the case of photons, online ART has become available in clinic with the introduction of MR-Linacs and the CBCT-based Ethos System by Varian. Although these machines provide the possibility of online ART, the processes are still fairly time consuming and require an approval of the adapted treatment plan by a clinician and a medical physicist. Since the initial treatment plan remains unchanged, the dose-guided patient positioning workflow evaluated in the scope of this habilitation does not require these approvals and could serve as an intermediate adaption step, ensuring an optimized reliable workflow. In cases where the dose-guided positioning is not sufficient, the online ART workflow can still be triggered using the same images.

A fast and reliable conversion of MR images to pseudoCTs is crucial for dose calculation on MR images. The presented Unet based conversion is hence of interest for clinics equipped with MR-Linacs as well as those using MR only workflows. Future research will be focused on a broader applicability in terms of further tumor locations as well as MR imaging sequences e.g. those used in clinical routine at MR-Linacs. Currently, due to technical reasons and the often smaller B-fields, the fast rt-4DMRI exploited in section 2.3.2. cannot be acquired on the MR-Linac machines. MR-Linacs though offer the possibility to provide tumor tracking based on online available cine slices acquired with a comparably high temporal resolution of e.g. 8 Hz in combination with gating of the beam delivery. The developed POP method could be used to optimize gating windows and hence to reduce overall treatment time per fraction.

In the case of proton therapy, image guided (online) adaption could be based on intensity corrected CBCTs, acquired prior to the treatment. In the scope of this habilitation a necessary QA procedure for (online) ART based on these images has been demonstrated. Future development will need to provide the necessary improvements in terms of speed and reliability of the workflow. Measures of artificial intelligence for CBCT intensity correction are currently exploited in this respect.

To conclude, the research conducted in the scope of this habilitation addressed various aspects of ART and uncertainty and sensitivity measures to further improve treatment planning and treatment plan adaption in radiotherapy with photons, protons and carbon ions paving the way towards a potential clinical adoption of the presented new techniques.



## 4 List of abbreviations

ART	Adaptive radiotherapy
CBCT	Cone-beam computed tomography
CERR	Computational environment for radiotherapy research
CT	Computed tomography
CTV	Clinical target volume
DIR	Deformable image registration
DSC	Dice similarity coefficient
DVH	Dose volume histogram
GPU	Graphics processing unit
GTV	Gross tumor volume
HU	Hounsfield unit
IGRT	Image-guided radiotherapy
IOV	Inter-observer variability
ITV	Internal target volume
M(A)E	Mean (absolute) error
MR	Magnetic resonance
MR-Linac	Magnetic resonance-linear accelerator
midV	Mid-ventilation target volume
OAR	Organ-at-risk
pCT	Planning CT
POP	Probability-of-presence
PRE	precision $PRE = (PTV_r / ITV_n^{5\%}) / PTV_r$
PTV	Planning target volume
RBE	Relative biological effectiveness
RMF	Repair-misrepair-fixation
RT	Radiotherapy
rt-4DMRI	Real-time 4D magnetic resonance imaging
RWD	RBE-weighted dose
RW-DVH	RBE-weighted dose volume histogram
SA	Sensitivity analysis
SE	Sensitivity $SE = (PTV_r / ITV_n^{5\%}) / ITV_n^{5\%}$
STAPLE	Simultaneous truth and performance level estimation
UA	Uncertainty analysis
Unet	U-shaped convolutional neural network
vCT	Virtual CT

## 5 References

- Acharya S *et al* 2016 Online Magnetic Resonance Image Guided Adaptive Radiation Therapy: First Clinical Applications *Int J Radiat Oncol Biol Phys* **94** 394–403
- Arbor N, Dauvergne D, Dedes G, Létang J M, Parodi K, Quiñones C T, Testa E and Rit S 2015 Monte Carlo comparison of x-ray and proton CT for range calculations of proton therapy beams *Phys. Med. Biol.* **60** 7585–99
- Carlson D J, Stewart R D, Semenenko V A and Sandison G A 2008 Combined use of Monte Carlo DNA damage simulations and deterministic repair models to examine putative mechanisms of cell killing *Radiation research* **169** 447–59
- Deasy J O, Blanco A I and Clark V H 2003 CERR: A computational environment for radiotherapy research *Med. Phys.* **30** 979–85
- Dinkla A M, Wolterink J M, Maspero M, Savenije M H F, Verhoeff J J C, Seravalli E, Išgum I, Seevinck P R and van den Berg C A T 2018 MR-Only Brain Radiation Therapy: Dosimetric Evaluation of Synthetic CTs Generated by a Dilated Convolutional Neural Network *Int J Radiat Oncol Biol Phys* **102** 801–12
- Eminowicz G, Rompokos V, Stacey C and McCormack M 2016 The dosimetric impact of target volume delineation variation for cervical cancer radiotherapy *Radiotherapy and oncology* **120** 493–9
- Han X 2017 MR-based synthetic CT generation using a deep convolutional neural network method *Med. Phys.* **44** 1408–19
- Hellebust T P, Tanderup K, Lervåg C, Fidarova E, Berger D, Malinen E, Pötter R and Petrič P 2013 Dosimetric impact of interobserver variability in MRI-based delineation for cervical cancer brachytherapy *Radiotherapy and oncology* **107** 13–9
- Hodapp N 2012 Der ICRU-Report 83: Verordnung, Dokumentation und Kommunikation der fluenzmodulierten Photonenstrahlentherapie (IMRT) *Strahlentherapie und Onkologie* **188** 97–9
- Jin C-B, Kim H, Liu M, Jung W, Joo S, Park E, Ahn Y S, Han I H, Lee J I and Cui X 2019 Deep CT to MR Synthesis Using Paired and Unpaired Data *Sensors (Basel, Switzerland)* **19**
- Lagendijk J J W, Raaymakers B W, van den Berg C A T, Moerland M A, Philippens M E and van Vulpen M 2014a MR guidance in radiotherapy *Phys. Med. Biol.* **59** R349-69
- Lagendijk J J W, Raaymakers B W and van Vulpen M 2014b The magnetic resonance imaging-linac system *Seminars in radiation oncology* **24** 207–9
- Lobefalo F *et al* 2013 Dosimetric impact of inter-observer variability for 3D conformal radiotherapy and volumetric modulated arc therapy: The rectal tumor target definition case *Radiation oncology (London, England)* **8** 176
- Lomax A J 2008 Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: The potential effects of inter-fraction and inter-field motions *Phys. Med. Biol.* **53** 1043–56
- Mutic S and Dempsey J F 2014 The ViewRay system: Magnetic resonance-guided and controlled radiotherapy *Seminars in radiation oncology* **24** 196–9
- Paganetti H 2014 Relative biological effectiveness (RBE) values for proton beam therapy. Variations as a function of biological endpoint, dose, and linear energy transfer *Phys. Med. Biol.* **59** R419-72

- Park Y-K, Sharp G C, Phillips J and Winey B A 2015 Proton dose calculation on scatter-corrected CBCT image: Feasibility study for adaptive proton therapy *Med. Phys.* **42** 4449–59
- Saltelli A 2007 *Sensitivity analysis of scientific models* (Hoboken, N.J., Chichester: Wiley; John Wiley [distributor])
- Schell S and Wilkens J J 2010 Advanced treatment planning methods for efficient radiation therapy with laser accelerated proton and ion beams *Med. Phys.* **37** 5330–40
- Scholz M, Kellerer A M, Kraft-Weyrather W and Kraft G 1997 Computation of cell survival in heavy ion beams for therapy *Radiat Environ Biophys* **36** 59–66
- Schuemann J, Dowdell S, Grassberger C, Min C H and Paganetti H 2014 Site-specific range uncertainties caused by dose calculation algorithms for proton therapy *Phys. Med. Biol.* **59** 4007–31
- Unkelbach J *et al* 2018 Robust radiotherapy planning *Phys. Med. Biol.* **63** 22TR02
- van Elmpt W, Landry G, Das M and Verhaegen F 2016 Dual energy CT in radiotherapy: Current applications and future outlook *Radiotherapy and oncology* **119** 137–44
- van Herk M 2004 Errors and margins in radiotherapy *Seminars in radiation oncology* **14** 52–64
- Vinod S K, Jameson M G, Min M and Holloway L C 2016 Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies *Radiotherapy and oncology* **121** 169–79
- Warfield S K, Zou K H and Wells W M 2004 Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation *IEEE transactions on medical imaging* **23** 903–21
- Wedenberg M, Lind B K and Hårdemark B 2013 A model for the relative biological effectiveness of protons: The tissue specific parameter  $\alpha/\beta$  of photons is a predictor for the sensitivity to LET changes *Acta oncologica (Stockholm, Sweden)* **52** 580–8
- Wolthaus J W H, Schneider C, Sonke J-J, van Herk M, Belderbos J S A, Rossi M M G, Lebesque J V and Damen E M F 2006 Mid-ventilation CT scan construction from four-dimensional respiration-correlated CT scans for radiotherapy planning of lung cancer patients *Int J Radiat Oncol Biol Phys* **65** 1560–71
- Yang M, Zhu X R, Park P C, Titt U, Mohan R, Virshup G, Clayton J E and Dong L 2012 Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration *Phys. Med. Biol.* **57** 4095–115

## 6 Acknowledgements

First of all, I would like to thank Prof. Claus Belka for the opportunity to pursue my habilitation at the LMU Department of Radiation Oncology under his supervision. I want to thank Prof. Katia Parodi and Prof. Peter Bartenstein for their participation in the Fachmentorat.

The prescribed research projects were performed during my time at the LMU Department of radiation oncology. I thus want to thank Prof. Belka for providing a very fruitful research environment and spirit and for his continuous support.

The projects would not have been possible without the help and support of many colleagues. Here I want to especially thank Sebastian Neppi, Jan Hofmaier, Moritz Rabe, Henning Schmitz, Asmus von Münchow, Michael Reiner and Christian Thieke. I am grateful for all the projects carried out with two outstanding colleagues Christopher Kurz and Guillaume Landry. Thanks for all the discussions, support and ideas throughout all the years. In addition I want to thank the entire Medical Physics team at the hospital for making daily work so cheerful and for all the laughs we shared during coffee breaks.

Many projects were a collaborative effort, mostly with the LMU Department of Medical Physics. Here I want to especially thank Katia Parodi, George Dedes and Chiara Gianoli for the very fruitful collaboration. In addition I am grateful for continuous support by David Carlson and Jan Wilkens.

Last but not least I want to thank my family for the enormous support and for making life besides work so joyful.

## 7 Declaration

Hiermit versichere ich an Eides statt, dass ich meine Habilitationsleistung selbständig und ohne andere als die angegebenen Hilfsmittel angefertigt habe, zudem die Herkunft des verwendeten und zitierten Materials ordnungsgemäß kenntlich gemacht habe.

Des Weiteren erkläre ich, dass ich mich weder anderweitig habilitiert, noch bereits Habilitationsversuche unternommen habe und dass mir kein akademischer Grad entzogen wurde oder ein Verfahren gegen mich anhängig ist, welches zur Entziehung eines akademischen Grades führen könnte.

Köln, 08. Januar 2023

Dr. rer. nat. Florian Kamp

## 8 Facsimile of relevant scientific contributions

# Variance-based sensitivity analysis for uncertainties in proton therapy: A framework to assess the effect of simultaneous uncertainties in range, positioning, and RBE model predictions on RBE-weighted dose distributions

Jan Hofmaier<sup>a)</sup>

*Department of Radiation Oncology, University Hospital, LMU Munich, Munich 81377, Germany*

George Dedes

*Department of Medical Physics, Faculty of Physics, LMU Munich, Garching b. Munich 85748, Germany*

David J. Carlson

*Department of Radiation Oncology, University of Pennsylvania, Philadelphia, PA 19104, USA*

Katia Parodi

*Department of Medical Physics, Faculty of Physics, LMU Munich, Garching b. Munich 85748, Germany*

Claus Belka

*Department of Radiation Oncology, University Hospital, LMU Munich, Munich 81377, Germany  
German Cancer Consortium (DKTK), Munich 81377, Germany*

Florian Kamp

*Department of Radiation Oncology, University Hospital, LMU Munich, Munich 81377, Germany*

(Received 12 August 2019; revised 20 October 2020; accepted for publication 11 November 2020; published 18 December 2020)

**Purpose:** Treatment plans in proton therapy are more sensitive to uncertainties than in conventional photon therapy. In addition to setup uncertainties, proton therapy is affected by uncertainties in proton range and relative biological effectiveness (RBE). While to date a constant RBE of 1.1 is commonly assumed, the actual RBE is known to increase toward the distal end of the spread-out Bragg peak. Several models for variable RBE predictions exist. We present a framework to evaluate the combined impact and interactions of setup, range, and RBE uncertainties in a comprehensive, variance-based sensitivity analysis (SA).

**Material and methods:** The variance-based SA requires a large number ( $10^4$ – $10^5$ ) of RBE-weighted dose (RWD) calculations. Based on a particle therapy extension of the research treatment planning system CERR we implemented a fast, graphics processing unit (GPU) accelerated pencil beam modeling of patient and range shifts. For RBE predictions, two biological models were included: The mechanistic repair-misrepair-fixation (RMF) model and the phenomenological Wedenberg model. All input parameters (patient position, proton range, RBE model parameters) are sampled simultaneously within their assumed probability distributions. Statistical formalisms rank the input parameters according to their influence on the overall uncertainty of RBE-weighted dose–volume histogram (RW-DVH) quantiles and the RWD in every voxel, resulting in relative, normalized sensitivity indices ( $S = 0$ : noninfluential input,  $S = 1$ : only influential input). Results are visualized as RW-DVHs with error bars and sensitivity maps.

**Results and conclusions:** The approach is demonstrated for two representative brain tumor cases and a prostate case. The full SA including  $\sim 3 \times 10^4$  RWD calculations took 39, 11, and 55 min, respectively. Range uncertainty was an important contribution to overall uncertainty at the distal end of the target, while the relatively smaller uncertainty inside the target was governed by biological uncertainties. Consequently, the uncertainty of the RW-DVH quantile  $D_{98}$  for the target was governed by range uncertainty while the uncertainty of the mean target dose was dominated by the biological parameters. The SA framework is a powerful and flexible tool to evaluate uncertainty in RWD distributions and DVH quantiles, taking into account physical and RBE uncertainties and their interactions. The additional information might help to prioritize research efforts to reduce physical and RBE uncertainties and could also have implications for future approaches to biologically robust planning and optimization. © 2020 The Authors. *Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine. [<https://doi.org/10.1002/mp.14596>]

**Key words:** proton therapy, range uncertainty, relative biological effectiveness, sensitivity analysis, uncertainty analysis

## 1. INTRODUCTION

Treatment plans in proton therapy are more prone to uncertainties than in photon therapy. In addition to setup uncertainty, which is also relevant for treatment with photons, proton beams are affected by range uncertainties. Furthermore, the relative biological effectiveness (RBE) of proton beams needs to be taken into account and additional sources of uncertainty are introduced through the conversion of physical to RBE-weighted dose (RWD). Since uncertainties in the actually delivered RWD may give rise to unexpected normal tissue toxicities or local treatment failure and may impede the intercomparability of different radiation modalities (photons, protons, heavier ions) in clinical studies, a well-founded understanding of the magnitude of the overall uncertainty and the impact and interactions of the different sources of uncertainty is crucial. To enable the comparison of different planning strategies (e.g., beam arrangements) with regard to these quantities, a systematic way to estimate them for an individual treatment plan is desirable. While there are many studies assessing the impact of physical<sup>1-4</sup> or biological uncertainties,<sup>5-7</sup> no method for the systematic assessment of the combined impact and interactions of setup, range, and biological uncertainties has been presented so far. When multiple sources of uncertainty are combined, the analysis is typically restricted to range and motion.<sup>8,9</sup> A possible way to deal with these uncertainties in intensity-modulated proton therapy (IMPT) is robust optimization. Up to now, robustness approaches are also mostly restricted to setup and range uncertainties and do not explicitly consider biological modeling. If at all, linear energy transfer (LET) is considered as a surrogate of biological effect.<sup>10,11</sup> Therefore, a better understanding of uncertainty in biological modeling in combination with setup and range uncertainties is needed to enable biologically robust planning. Furthermore, quantifying the relative impact of setup, range and RBE uncertainty on the overall uncertainty of clinically relevant dose metrics could help to prioritize research efforts aiming at reducing the individual uncertainties and improve cost-effectiveness in radiotherapy. The technique of global, variance-based sensitivity analysis (SA) is a method to evaluate the influence of the uncertainty in various input factors on the output of a quantitative model.<sup>12</sup> Compared to local methods, such as derivative-based approaches, regression analysis, or the isolated treatment of the different input factors (one factor at a time approaches), this technique has the advantage of exploring the entire input space by varying all input factors simultaneously, which allows to take into account also interactions between multiple input factors.<sup>13</sup> While the alternative techniques mentioned have their limitations in the case of nonlinear models, global, variance-based SA is a model-independent approach and is applicable for any probability distribution of the input factors. In the field of medical physics, this technique has so far only been applied to RBE modeling

of carbon ion therapy, excluding range and setup uncertainties<sup>14,15</sup> and to nuclear medicine, in order to assess the impact of interpatient variability on organ dose estimates.<sup>16</sup> In this feasibility study, we present a framework to apply the technique of global, variance-based SA to uncertainties in proton therapy, explicitly modeling RBE, range, and setup errors.

## 2. MATERIALS AND METHODS

### 2.A. Global variance-based sensitivity analysis

Global variance-based SA is a method to estimate the relative influence of the  $k$  input factors  $X = (x_1, x_2, \dots, x_k)$  on the output  $Y$  of a model  $f$ :

$$Y = f(X) \quad (1)$$

The function  $f$  can be decomposed into terms of increasing dimensionality whose mean is zero, that is<sup>17</sup>

$$f = f_0 + \sum_l^k f_l(x_l) + \sum_{l=1}^k \sum_{m>l}^k f_{lm}(x_l, x_m) + \dots + f_{12\dots k}(x_1, x_2, \dots, x_k) \quad (2)$$

where for all  $p = 1, \dots, s$

$$\int f_{1\dots s}(x_1, \dots, x_s) dx_p = 0 \quad (3)$$

Sobol' proved that then all summands in equation (2) are orthogonal.<sup>17</sup> The variance in  $Y$  can be decomposed<sup>12</sup>

$$V(Y) = \sum_{l=1}^k V_l + \sum_{l=1}^k \sum_{m>l}^k V_{lm} + \dots + V_{1\dots k} \quad (4)$$

where

$$V_l = V(f_l) = V[E(Y|X_l)] \quad (5)$$

The expectation value  $E(Y|X_l)$  is hereby calculated over all possible values of all input factors except for  $X_l$ , which is kept fixed. The higher order terms are

$$V_{lm} = V(f_{lm}) = V[E(Y|X_l, X_m)] - V_l - V_m \quad (6)$$

and so on. The first- and second-order sensitivity indices introduced by Sobol' are defined as<sup>12</sup>:

$$S_l = \frac{V_l}{V(Y)} \quad (7)$$

$$S_{lm} = \frac{V_{lm}}{V(Y)} \quad (8)$$

Higher order terms are defined in an analogous fashion. Monte Carlo estimates for all sensitivity indices can be calculated without the need to know an explicit form of  $f$  or any of the terms in the expansion in Eq. (2). Due to the normalization to the overall variance, the Sobol' indices are normalized to 1. Since the number of sensitivity indices is  $(2^k - 1)$  for  $k$  input factors, making interpretation of the results very difficult, total effects  $ST_l$  are introduced.<sup>18</sup> For the input factor  $l$  they are defined as the sum of all terms of any order containing  $l$ :



$$ST_l = S_l + \sum_{m \neq l}^k S_{lm} + \dots + S_{1,\dots,k} \quad (9)$$

when only first order and total effects for each input factor are considered, the number of indices is reduced to  $2k$ . First order and total effects allow for an intuitive interpretation:  $S_l$  is the average fraction by which the overall variance would be reduced if input factor  $l$  could be fixed anywhere in its range,  $ST_l$  is the average fraction of the overall variance that would remain if all input factors except for  $l$  could be fixed within their respective range.  $ST_l = 0$  is the necessary and sufficient condition for input  $l$  being noninfluential. By examining the difference,  $(ST_l - S_l)$ , the impact of interaction terms involving input factor  $l$  can be characterized. If  $(ST_l - S_l) = 0$ , interactions with input factor  $l$  do not contribute to overall variance.

Saltelli<sup>19</sup> proposed an efficient method for direct Monte Carlo calculation of  $S_l$  and  $ST_l$ , without the need to calculate all the interaction terms, which has also been used in this paper.

In this approach,  $S_l$  and  $ST_l$  are estimated via:

$$S_l = \frac{\frac{1}{N} \sum_{m=1}^N f(B)_m \cdot [f(A^{(l)})_m - f(A)_m]}{V(Y)} \quad (10)$$

$$ST_l = \frac{\frac{1}{2N} \sum_{m=1}^N [f(A)_m - f(A^{(l)})_m]^2}{V(Y)} \quad (11)$$

where  $A$  and  $B$  are independently sampled input matrices of  $N$  input vectors (size:  $N \times k$ ). The matrix  $A^{(l)}$  is equal to matrix  $A$ , except for column  $l$ , which is taken from  $B$ .  $(A)_m$  and  $(B)_m$  are the  $m$ -th rows of  $A$  and  $B$ , respectively. The total number of model evaluations in this approach is  $N \cdot (k + 2)$ .  $N$  has to be chosen sufficiently large for Eqs. (10) and (11) to converge.

A faster convergence of Eqs. (10) and (11) is achieved when the input parameters are sampled from quasi-random, low-discrepancy sequences.<sup>19</sup> In our implementation, we used the Sobol' sequence<sup>20,21</sup> as suggested by Saltelli.

In our application of the concept, the model  $f$  will correspond to an RWD distribution calculation including a calculation of RBE-weighted dose volume histograms (RW-DVHs), the input vector  $X$  will contain isocenter shifts in three spatial dimensions, relative and absolute range shifts as well as biological model parameters. The output  $Y$  will include the dose in every voxel and RW-DVH quantiles for structures of interest. Since for the variance-based SA the model needs to be evaluated approximately  $10^4$ – $10^5$  times, a fast RWD calculation for any set of  $X$  from the input space is required. To achieve this, a GPU-based RWD calculation was implemented based on a particle extension of the research treatment planning system CERR.<sup>22–26</sup> To model the physical uncertainties, the following approximations were made: The proton beams were assumed to be nondivergent, and patient deformations and rotations were excluded. Range uncertainty was modeled as a relative and an absolute range shift, which was applied to all spots of the same beam equally. A detailed

description of the implementation can be found in Appendices A and B.

In clinical routine, a constant RBE of 1.1 is commonly assumed. However, there is evidence from *in vitro* experiments that RBE is dependent on dose, biological endpoint, and proton energy and there is an ongoing debate if the current clinical practice needs to be revised.<sup>5–7,27,28</sup> For variable RBE prediction, two biological models are currently implemented: the mechanistic repair-misrepair-fixation (RMF) model,<sup>29,30</sup> which uses double strand break (DSB) yields from a Monte Carlo Damage simulation (MCDS)<sup>31</sup> and the phenomenological Wedenberg model.<sup>32</sup> Both models provide a method to calculate radiosensitivity parameters of the linear quadratic (LQ) model,  $\alpha_P$  and  $\beta_P$  and have the advantage that they can be executed very fast for changed model and x-ray reference radiosensitivity parameters. For the RMF model, the DSB yield  $\Sigma$  and the x-ray reference parameters  $\alpha_X/\beta_X$  were treated as uncertain; details on the implementation can be found in Appendix C.1. For the Wedenberg model, x-ray reference parameters  $\alpha_X/\beta_X$ , the fit parameter  $q$  and the model assumption  $\beta_P = \beta_X$  were treated as uncertain. Details on the implementation can be found in Appendix C.2.

## 2.B. Application to patient cases

The framework was applied to two brain tumor patient cases. The evaluation of an additional prostate case can be found in the supplementary material S1. In both brain tumor cases the clinical target volume (CTV) was overlapping with the optic nerve and in close proximity to the brain stem. For patient 1, the CTV partially overlapped with the optic chiasm while for patient 2 the chiasm was almost completely within the CTV. The CTV of patient 1 had a larger volume with a size of 15.2 cm<sup>3</sup>, for patient 2 it was 4.7 cm<sup>3</sup>. Planning target volumes were created using an isotropic CTV-to-PTV margin of 3 mm. The plan for patient 1 consisted of two PBS beams from 60° and 135°, the plan for patient 2 of two opposing PBS beams from 90° and 270°. Both plans were optimized for a fraction RWD of 1.8 Gy (RBE), where a constant RBE of 1.1 was assumed. The total prescribed RWD was 54 Gy (RBE) in both cases. In the plan optimization, the total  $D_{2\%}$  for adjacent and overlapping OARs (chiasm, optic nerves, brain stem) was constrained to be smaller or equal to 54 Gy (RBE). In each plan, the two beams were optimized independently to deliver a homogeneous dose distribution to the target (single field uniform dose (SFUD) concept). Sensitivity analyses of the resulting plans were performed using the RMF and the Wedenberg model. All input parameter uncertainties were assumed to follow normal distributions truncated to two standard deviations ( $\sigma$ ). The following  $\sigma$  were used:

- 1 mm for patient shifts in X, Y, and Z direction<sup>33</sup>
- 3% for relative range uncertainty<sup>1</sup> ( $R_{rel}$ )
- 1 mm absolute range uncertainty<sup>1</sup> ( $R1_{abs}$ ,  $R2_{abs}$  for beam numbers 1,2,...)

- 10% for the x-ray reference radiosensitivity parameters ( $\alpha_X/\beta_X$ )
- 15% for the parameter  $q$  of the Wedenberg model<sup>32</sup>
- 10% for  $\beta_p$  in the Wedenberg model
- 5% for the DSB yield  $\Sigma$  used in the RMF model

These assumptions are not definitive and might differ between tumor sites (e.g., in the abdomen a larger setup error than 1 mm might be adequate) or CT acquisition (e.g., in case a dual-energy CT is used for stopping power determination a smaller relative range uncertainty might be reasonable<sup>34</sup>). To model a possible higher uncertainty in the x-ray reference sensitivity parameters in organs at risk (e.g., the chiasm), an additional calculation was performed where  $\alpha_X/\beta_X$  was assumed to be uniformly distributed over the interval 1.5 to 10 Gy. All other input uncertainties remained unchanged. For the CTV and the chiasm, the overall uncertainty of all RW-DVH quantiles was visualized by plotting the RW-DVH with its respective 95% and 68% confidence intervals. For selected quantiles of (CTV  $D_{98\%}$ , CTV  $D_{50\%}$ , brain stem  $D_{2\%}$ , optic nerve  $D_{2\%}$ , and chiasm  $D_{2\%}$ ), first order and total effect sensitivity indices were calculated. Additionally, the first-order sensitivity for the RBE-weighted dose was calculated on a voxelwise basis and visualized as sensitivity map.

To demonstrate the application of the framework to a prostate case, an additional evaluation of a such a plan can be found in supplementary material S1.

### 3. RESULTS

A fast modeling of patient shifts, range shifts and changes in biological parameters was implemented which allows the calculation of RWD distributions for arbitrary sets of these input parameters from the input space. The simultaneous variation of all input parameters allows to model interactions of different sources of uncertainty and to perform a global, variance-based SA. Here the results for the two brain tumor cases are shown, the results for the prostate case can be found in the supplementary material S1.

On a computer with 16 CPU cores (Intel Xeon E5-2690 @ 2.90 GHz), 192 GB RAM and two Nvidia Tesla K80 GPUs the full SA including  $\sim 3 \times 10^4$  RWD calculations was performed in 39 min for patient 1 and 11 min for patient 2.

Figure 1 shows the convergence of the first-order indices and total effects for selected DVH quantiles, a representative voxel at the center of the PTV and a representative voxel in the high LET region for patient 1.  $N$  refers to the number of rows of the input matrices used for the Saltelli estimator for  $S_i$  and  $ST_i$  described above, the actual number of RWD calculations performed is  $N \cdot (k + 2)$ , with the number of input factors  $k$ .

#### 3.A. DVH quantiles

Figures 2 and 3 show the RW-DVH of the CTV and the chiasm for patients 1 and 2, respectively. The proton treatment plan was optimized on 1.8 Gy(RBE) in tumor assuming

a constant RBE of 1.1. Then the SA was performed with both biological models assuming a spatially constant  $\alpha_X/\beta_X = 2$  Gy ( $\alpha_X = 0.1$  Gy<sup>-1</sup> and  $\beta_X = 0.05$  Gy<sup>-2</sup>).<sup>35</sup> For each model, the treatment plans were recalculated  $\sim 3 \times 10^4$  times, randomly varying patient position, proton range and RBE model parameters within their assumed uncertainties. To quantify the overall uncertainty, 95% and 68% confidence intervals were calculated empirically from the resulting RW-DVHs and visualized in Figs. 2 and 3. As expected, a higher RBE than 1.1 is predicted for both biological models. The same calculation was in addition also applied to an RWD calculation assuming a constant RBE of 1.1., including only range and setup uncertainties. The resulting RW-DVH is plotted for comparison. For both patients, a larger overall uncertainty was observed for the variable RBE models with their respective uncertainties included. For example, the expectation value for the mean RWD to the CTV for patient 1 was  $2.04^{+0.19}_{-0.19}$  Gy(RBE),  $2.03^{+0.14}_{-0.14}$  Gy(RBE) and  $1.77^{+0.04}_{-0.11}$  Gy (RBE) for the RMF model, the Wedenberg model and a constant RBE of 1.1, respectively (the reported ranges are the 95% confidence intervals.). For patient 2, the mean RWD to the CTV was  $2.01^{+0.21}_{-0.22}$  Gy(RBE),  $2.00^{+0.16}_{-0.17}$  Gy(RBE) and  $1.75^{+0.04}_{-0.14}$  Gy(RBE) for the RMF model, the Wedenberg model and a constant RBE of 1.1, respectively. For the chiasm, the  $D_2$  for patient 1 was  $2.23^{+0.23}_{-0.23}$  Gy(RBE) and  $2.17^{+0.18}_{-0.18}$  for the RMF model and the Wedenberg model, respectively, when an  $\alpha_X/\beta_X = 2$  Gy with a standard deviation of 10% was assumed. In the calculation with the large  $\alpha_X/\beta_X$  variation it was  $2.09^{+0.28}_{-0.23}$  Gy(RBE) (RMF) and  $1.98^{+0.28}_{-0.20}$  Gy(RBE) (Wedenberg). For patient 2, a chiasm  $D_2$  of  $2.12^{+0.25}_{-0.24}$  Gy(RBE) and  $2.09^{+0.19}_{-0.17}$  Gy(RBE) was observed for RMF and Wedenberg, respectively, when an  $\alpha_X/\beta_X = 2$  Gy with a standard deviation of 10% was assumed.  $2.01^{+0.26}_{-0.24}$  Gy(RBE) (RMF) and  $1.93^{+0.24}_{-0.16}$  Gy(RBE) (Wedenberg) was found for the  $D_2$  in the calculation with the large  $\alpha_X/\beta_X$  variation.

For selected clinically relevant RW-DVH quantiles, the confidence intervals in Figs. 2 and 3 are broken down to the impact of the different uncertainties in terms of  $S_i$  and  $ST_i$  in Fig. 4. All plots show results of the calculation with  $\alpha_X/\beta_X = 2$  Gy and a standard deviation of 10%. SA results are color-coded for both patients and both RBE models. For most quantiles, the differences between the  $S_i$  and  $ST_i$  are small, indicating a low impact of interaction terms on the overall uncertainty. For range uncertainty, however, interactions often do play a role. For example, for the  $D_2$  of the right optic nerve for patient 2 [Figs. 4(c) and 4(d)]  $ST_i$  is considerably larger than  $S_i$  both for the relative range uncertainty ( $R_{rel}$ ) for both biological models ( $ST_{R_{rel}} = 0.56$ ,  $S_{R_{rel}} = 0.26$  for the RMF model and  $ST_{R_{rel}} = 0.54$ ,  $S_{R_{rel}} = 0.28$  for the Wedenberg model, respectively) and for shifts in  $Y$  and  $Z$  direction ( $S_Y = 0.03$ ,  $S_Z = 0.18$ ,  $ST_Y = 0.25$ ,  $ST_Z = 0.45$  for the RMF model and  $S_Y = 0.07$ ,  $S_Z = 0.20$ ,  $ST_Y = 0.26$ ,  $ST_Z = 0.44$  for the Wedenberg model, respectively). This suggests that, in this plan, a relevant fraction of the overall  $D_2$  uncertainty for the right optic nerve is attributable to interaction between setup and range uncertainty. Biological

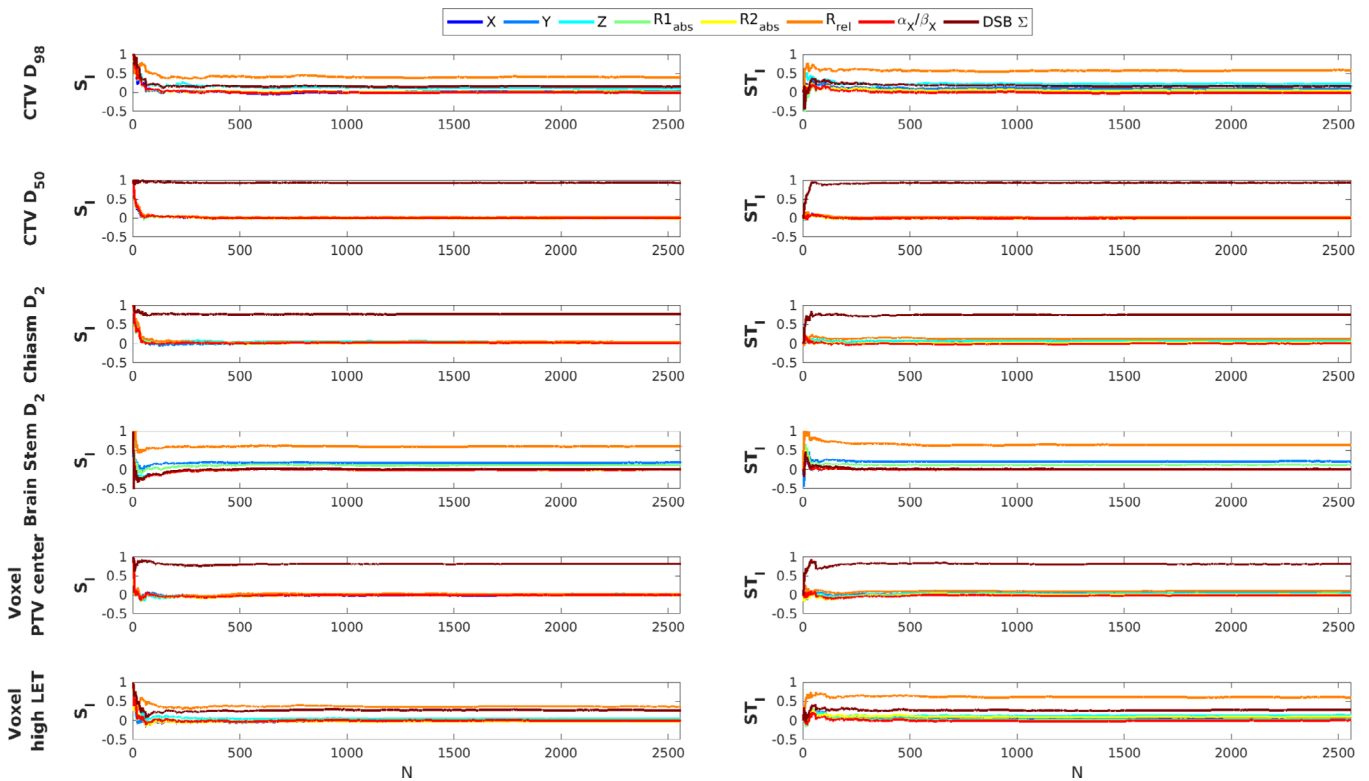


FIG. 1. Convergence of first-order sensitivity indices and total effects with sample size  $N$  for selected RW-DVH quantiles and two representative voxels for patient 1. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

uncertainty generally was driven by  $DSB\Sigma$  for the RMF model and by  $q$  and  $\beta_p$  for the Wedenberg model and the most important contribution to the CTV  $D_{50}$  (e.g., for patient 2:  $S_{DSB\Sigma} = 0.75$ ,  $ST_{DSB\Sigma} = 0.77$  for the RMF model and  $S_q = 0.42$ ,  $S_{\beta_p} = 0.23$ ,  $ST_q = 0.43$ ,  $ST_{\beta_p} = 0.23$  for the Wedenberg model, respectively). In both models, the relative impact of the x-ray reference parameters  $\alpha_x/\beta_x$  was very low in comparison (e.g., for the CTV  $D_{50}$  for patient 2:  $S_{(\alpha/\beta)_x} < 0.01$ ,  $ST_{(\alpha/\beta)_x} < 0.01$  for the RMF model and  $S_{(\alpha/\beta)_x} = 0.02$ ,  $ST_{(\alpha/\beta)_x} = 0.02$  the Wedenberg model, respectively). Biological uncertainty was also the most important contribution to the chiasm  $D_2$  for patient 1 ( $S_{DSB\Sigma} = 0.78$  for the RMF model and  $S_q = 0.41$  and  $S_{\beta_p} = 0.14$  for the Wedenberg model). The relative range uncertainty  $R_{rel}$  was an important input factor for many investigated DHV quantiles with the exception of the  $D_{50}$  of the CTV and the  $D_2$  of the brain stem for patient two, which due to its position lateral to the two opposing beams from  $90^\circ$  and  $270^\circ$  was not affected by range shifts. For this parameter, the most relevant contribution to overall uncertainty is observable for a patient shift in  $Y$  direction ( $S_Y = 0.71$  for the RMF model and  $S_Y = 0.69$  for the Wedenberg model, respectively).

### 3.B. Voxelwise SA

The result of the voxel-based SA assuming  $\alpha_x/\beta_x = 2$  Gy with a standard deviation of 10 % for patient 1 is shown in Fig. 5 for the RMF model and in Fig. 6 for the Wedenberg model. Nominal RWD

distribution, the local standard deviation as a measure of local uncertainty and the dose-weighted LET distribution are shown in the first row. SA maps report the contribution of the input uncertainties to the local variance for every voxel, indicating the spatial changes of the impact of different uncertainties. The largest uncertainties are observed at the distal end of the beams, where they are governed by  $R_{rel}$ . The impact of the absolute range uncertainties  $R1_{abs}$  and  $R2_{abs}$  is small in comparison, as well as the uncertainty in the x-ray reference parameters  $\alpha_x/\beta_x$ . In the CTV and in the entrance plateaus, the biological input factors  $\beta_p$  for the Wedenberg and  $DSB\Sigma$  for the RMF model are the most important contributions, where the overall uncertainty is generally lower than at the distal ends of the beams.

## 4. DISCUSSION

The presented framework is, to the best of our knowledge, the first implementation of a tool for variance-based sensitivity analysis of the combined impact of setup, range and RBE uncertainties in proton therapy, including also their interactions. Additionally to the numerical calculation of confidence intervals for the RBE-weighted dose in every voxel and RW-DVH quantiles, it allows to break down the overall uncertainty to the impact of the different sources of uncertainty. This complimentary sensitivity information has not yet been reported or used in proton therapy. The computation times were below 40 min which is extremely fast given the recalculation of  $\sim 3 \times 10^4$

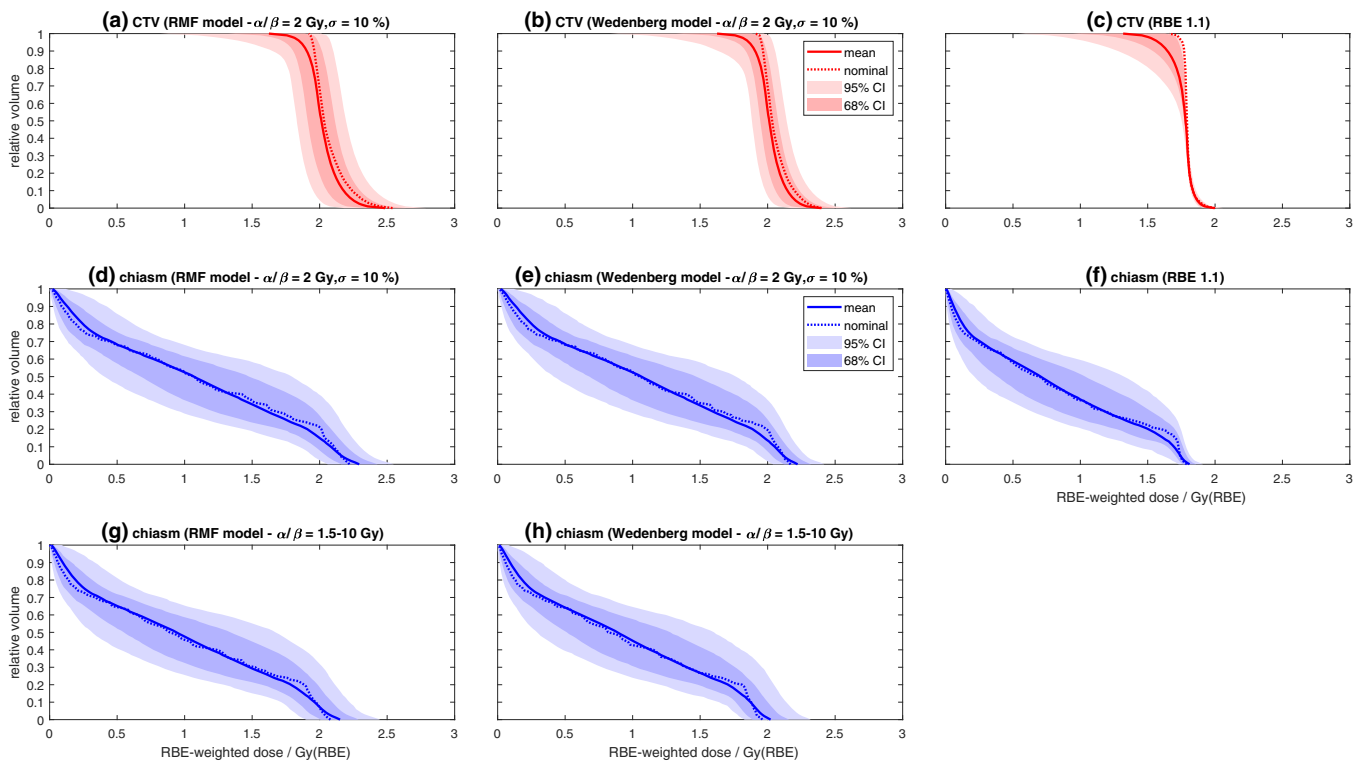


FIG. 2. RW-DVHs for the clinical target volume and the chiasm for a plan optimized for a constant RBE of 1.1 for patient 1. The uncertainty analysis was performed by recalculating the dose using the RMF (a and d) and the Wedenberg model (b and e) including range, setup, and RBE uncertainty. Panels (c) and (f) show the RW-DVHs for a constant RBE of 1.1 including only range and setup uncertainty. Panels (g) and (h) show the variation of the RW-DVH of the chiasm including range, setup and RBE uncertainty when  $\alpha_X/\beta_X$  is varied over the larger interval from 1.5 to 10 Gy for comparison. 68% and 95% confidence intervals of the RW-DVHs are visualized by the shaded areas. “Nominal” refers to a forward calculation in the respective model with all input factors fixed to their nominal value. The solid line shows the expectation value of the DVH. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

(corresponding to  $N = 2560$  in the Saltelli formalism) treatment plans per SA execution for patient 1. For patient 2, which has a considerably smaller CTV the simulation time was even shorter. This was expected, since both the number of treated voxels (due to the restriction to the 2 cm expansion of the target volume) and the number of PBS spots increase with the target volume and therefore the number of entries in the  $ij$ -matrices increase. The quick convergence of the sensitivity indices suggests that actually a much smaller number of about  $N = 500$  would already be sufficient. Since the calculation time scales linearly with the number of dose calculations, this would mean a reduction by 80%, to well below 10 minutes for patient 1 and to less than 2 and a half minutes for patient 2. The SA was performed for RW-DVH quantiles and the RWD in every voxel of interest. Further plan quality metrics based on the RWD distribution such as equivalent uniform dose, homogeneity index, conformity index, tumor control, and normal-tissue complication probabilities and others could be included at very little extra computational cost. An alternative approach to modeling uncertainties was proposed by Bangert *et al.*,<sup>8</sup> who introduced analytical probabilistic modeling (APM), a technique to propagate setup and range uncertainties through a pencil beam dose calculation via analytical integration to calculate expectation

values and variances for dose distribution and other plan quality indicators. Wieser *et al.*<sup>36</sup> used APM to investigate the influence of setup and range uncertainties on RWD distributions, however, uncertainties in the biological modeling itself were not considered in their work. Note that both works do not include the possibility to determine sensitivity values. Perkó *et al.*<sup>9</sup> used an alternative approach to sensitivity analysis using polynomial chaos expansion, but also did not consider uncertainties in RBE modeling.

#### 4.A. Potential applications

The current performance of the SA framework is sufficient to be forward calculated for the clinical evaluation of proton treatment plans. In such a setting, the additional uncertainty and sensitivity information could support the decision for or against a treatment plan and help to find the optimal compromise. Forward calculation of the SA could also be used in planning studies for the systematic assessment of the impact of setup, range, and RBE uncertainty on clinically relevant dosimetric parameters in proton therapy. The information which type of uncertainty is dominating the overall uncertainty could help to prioritize research attempts to reduce the uncertainties. In this regard, higher cost effectiveness could be achieved by concentrating on the dominant contributions



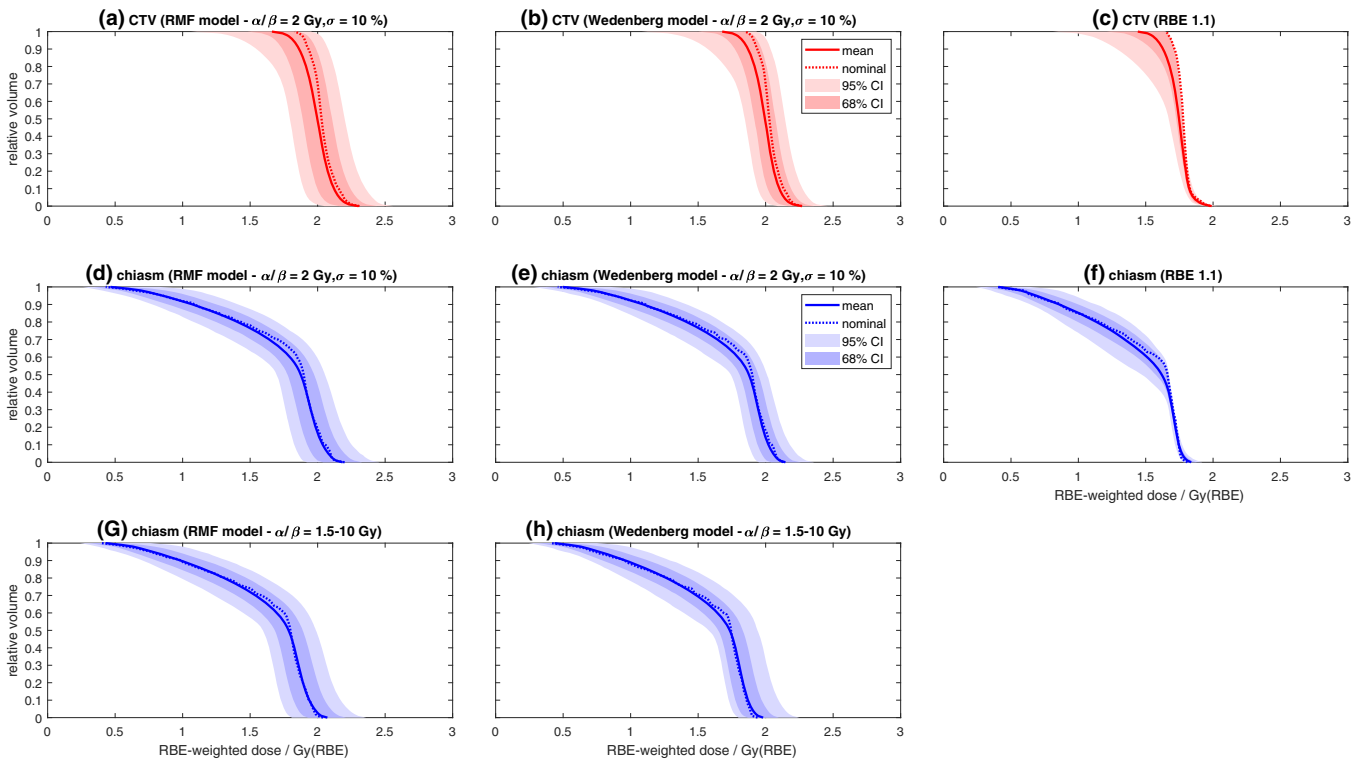


FIG. 3. RW-DVHs for the CTV and the chiasm for a plan optimized for a constant RBE of 1.1 for patient 2. The uncertainty analysis was performed by recalculating the dose using the RMF (a and d) and the Wedenberg model (b and e) including range, setup, and RBE uncertainty. Panels (c) and (F) show the RW-DVHs for a constant RBE of 1.1 including only range and setup uncertainty. Panels (g) and (h) show the variation of the RW-DVH of the chiasm including range, setup, and RBE uncertainty when  $\alpha_X/\beta_X$  is varied over the larger interval from 1.5 to 10 Gy for comparison. 68% and 95% confidence intervals of the RW-DVHs are visualized by the shaded areas. “Nominal” refers to a forward calculation in the respective model with all input factors fixed to their nominal value. The solid line shows the expectation value of the DVH. [Color figure can be viewed at wileyonlinelibrary.com]

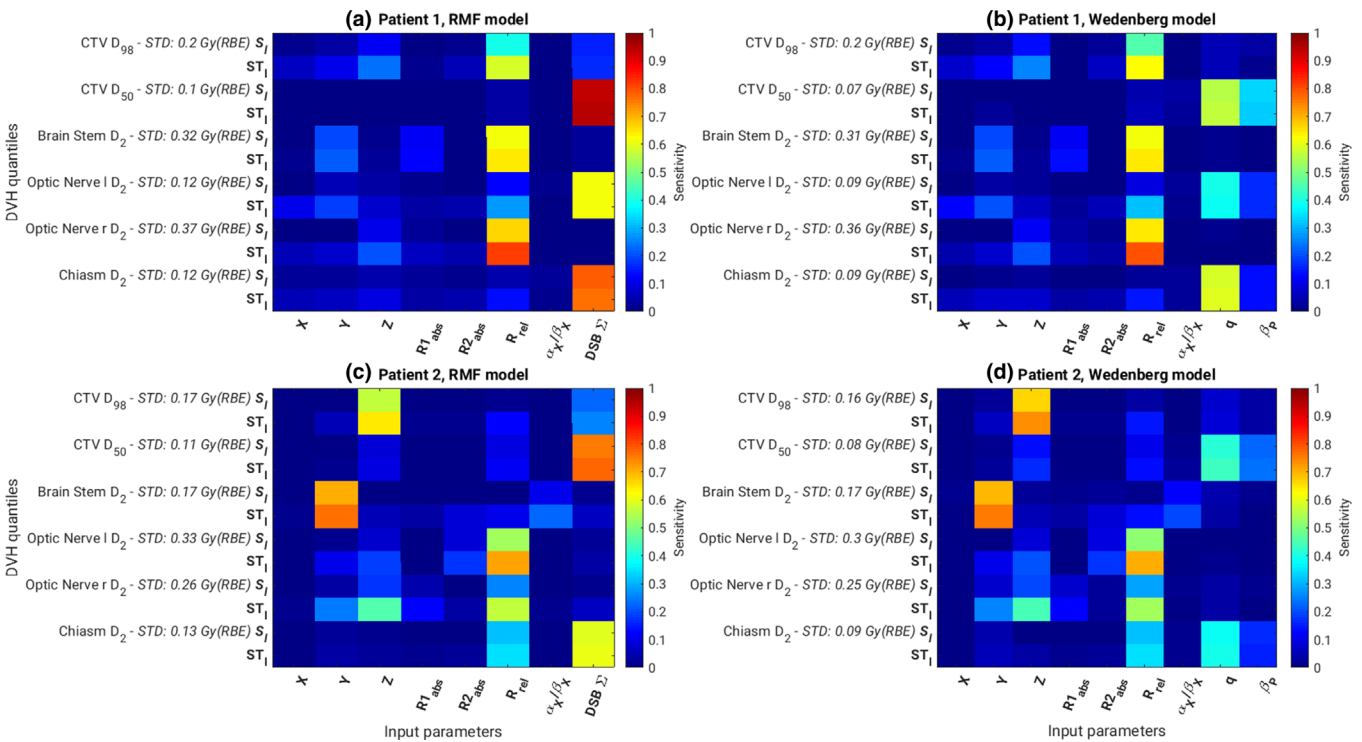


FIG. 4. First-order sensitivities and total effects for selected DVH quantiles of the CTV, brain stem, the optic nerves, and the chiasm for both patients and both RBE models in the calculation with  $\alpha_X/\beta_X = 2$  Gy with a standard deviation of 10%. The empirical standard deviations of the respective quantile are also reported. [Color figure can be viewed at wileyonlinelibrary.com]

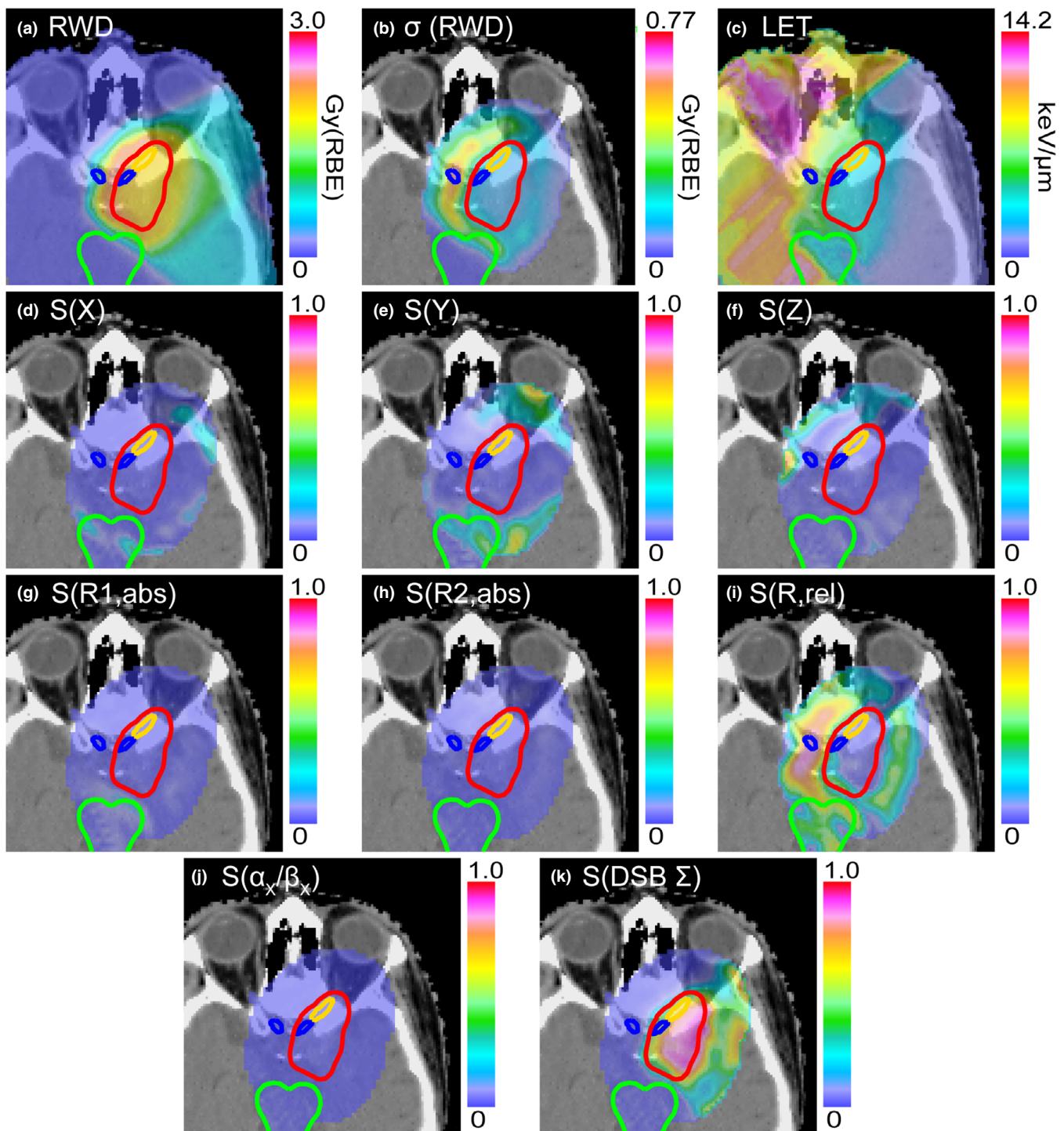


FIG. 5. RWD distribution, local standard deviation, LET distribution, and sensitivity maps for patient 1 and the RMF model in the calculation with range, setup, and RBE uncertainties included. For  $\alpha_x/\beta_x$ , a nominal value of 2 Gy and a standard deviation of 10% was assumed. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

to overall uncertainty. Another possible application of the SA framework could be the systematic comparison of proton treatment plans to evaluate, for example, different robust planning concepts.<sup>10,11</sup> This would allow to determine the residual uncertainty of RW-DVH quantiles of interest for these plans and analyze the sources of this uncertainty using the sensitivity indices.

#### 4.B. Limitations

Current limitations of our SA framework include the restriction to rigid patient shifts, excluding rotations. Deformations are not explicitly modeled, either, although some nonrigid changes (such as weight loss and filling of air cavities with fluid) are modeled by the employed heuristic model

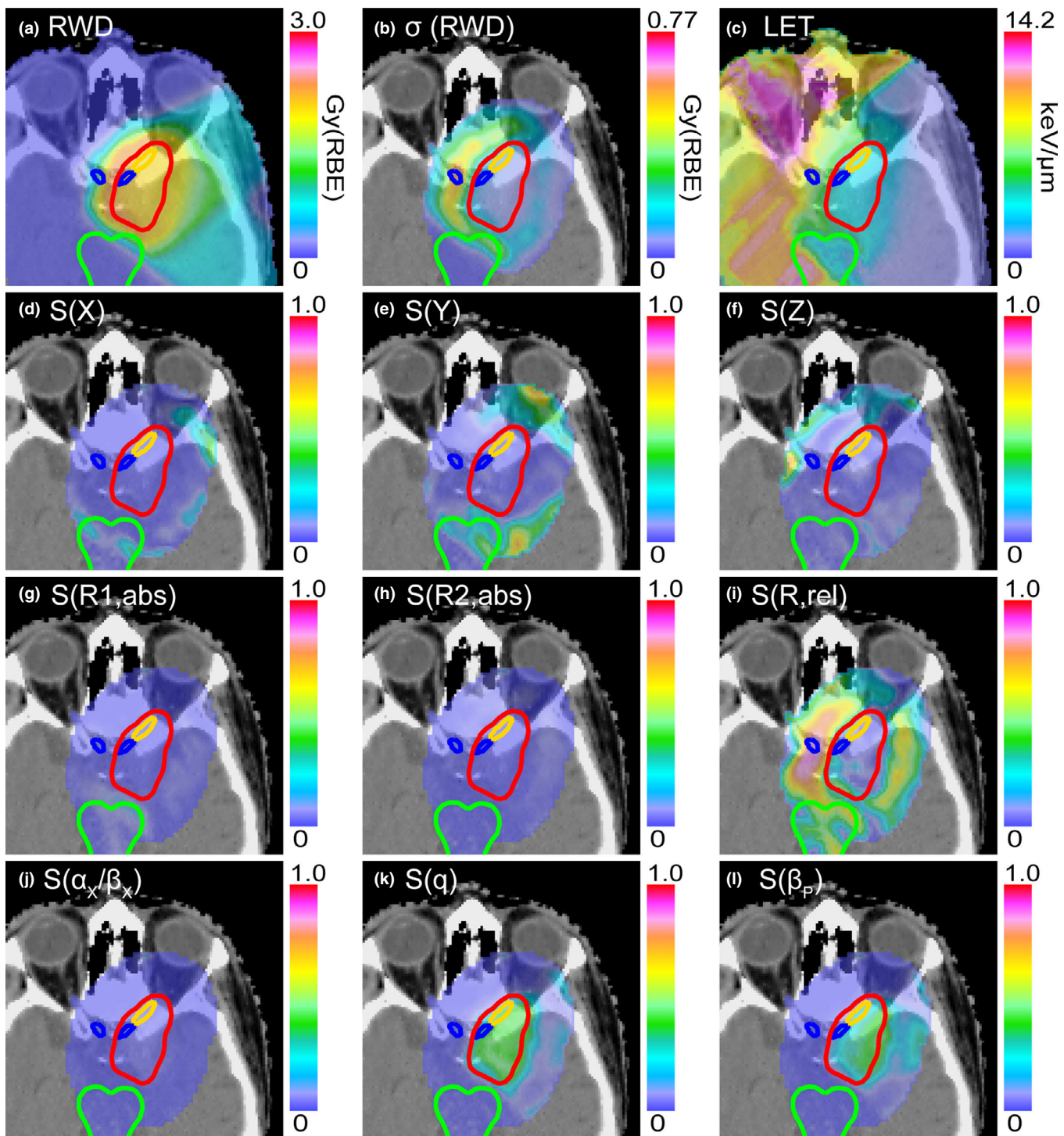


FIG. 6. RWD distribution, local standard deviation, LET distribution, and sensitivity maps for patient 1 and the Wedenberg model in the calculation with range, setup, and RBE uncertainties included. For  $\alpha_x/\beta_x$ , a nominal value of 2 Gy and a standard deviation of 10% was assumed. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

of absolute range uncertainties. However, considerable deformations of the patients' anatomy cannot be modeled. Furthermore, the method does not yet cover all possible types of uncertainties. Additional uncertainties exist in radiotherapy planning and delivery, which are not included in our approach, for example, inter- and intraobserver delineation variability. In this work, a pencil beam algorithm based on

precalculated Geant4 simulated data in water was used. Although Monte Carlo algorithms are known to be more accurate than pencil beam algorithms, the necessary high number of RWD calculations cannot be achieved with a Monte Carlo algorithm in reasonable time. The accuracy of pencil beam algorithms is known to decrease in regions with high tissue heterogeneity, therefore, results obtained with the



SA framework using the current pencil beam algorithm will be less reliable in such cases.

Having a good assumption of the underlying uncertainties in the input factors is crucial for the execution of the SA. Unfortunately, these uncertainties can be difficult to estimate, in particular for the biological parameters. For example, a 95% confidence interval of about  $\pm 18\%$  was reported for  $q$  as a fit parameter over various in-vitro experiments,<sup>32</sup> which corresponds to standard deviation of 9% for normal distributed data. To account for additional uncertainties related to the transfer from in-vitro data to the clinic, a standard deviation of 15% was assumed for  $q$ . The 5% standard deviation for DSB  $\Sigma$  in the RMF model was based on reasoning about the underlying Monte Carlo simulation of DSB induction.<sup>37</sup> However, these should be considered rather rough estimates of the actual uncertainty. Therefore, we did not aim at comparing the RBE prediction uncertainties of the two models. The objective is to show the flexibility of the approach with regard to the used biological model. It should also be kept in mind that the variance-based SA gives information about the variability, that is, the precision of the output, but does not give any information about the accuracy of a model. This means, for instance, that a higher robustness of an RBE model against uncertainties in its input parameters does not imply a higher accuracy.

In our analysis, the impact of the x-ray reference parameter  $\alpha_X/\beta_X$  on the overall RWD variability was low compared to the other biological parameters when the RWD of a single fraction was calculated, even when  $\alpha_X/\beta_X$  was varied over the large interval from 1.5 to 10 Gy. This means that the used RBE models are rather robust against  $\alpha_X/\beta_X$  variability. It should be stressed, however, that this does not include fractionation effects.

#### 4.C. Outlook

In this first analysis we applied the SA framework to proton therapy. The application to other charged particle types, such as helium or carbon ions could be achieved in the same way. Given that for these heavier ions the RBE is generally expected to be higher than for protons, a systematic assessment of uncertainty might be even more important in these cases.

Given the present performance of our SA framework, also future applications in robust plan optimization itself are imaginable if the execution of the variance-based SA can be further accelerated. While the framework is currently limited to the forward evaluation of proton treatment plans since the code still takes too long to be executed during optimization (for our patient cases from a few minutes to more than half an hour), the quick convergence of the sensitivity indices suggests that a considerable reduction of the number of model evaluations is feasible. In addition, the sensitivity analysis could be restricted to regions of interest with respect to plan robustness. The use of multiple, high-end GPUs with more memory for this highly parallelizable code is also expected to significantly improve the performance. All this might

accelerate the estimation of the sensitivity indices to a point where they can be evaluated during optimization. To date, most robust optimization approaches do not include RBE variability.<sup>10</sup> If at all, RBE is only considered indirectly using LET as a surrogate. While it is in theory possible to fully compensate for setup and range uncertainty (although at the cost of additional dose to normal tissue), this is not the case for RBE uncertainty. Although RBE uncertainty can be reduced by avoiding excessive LET hot spots, it cannot be fully eliminated by shaping the physical dose distribution. Explicit inclusion of RBE uncertainty would therefore lead to an inevitably larger overall uncertainty, rendering current robust optimization approaches insufficient. The complimentary sensitivity information has the potential to overcome these limitations since it allows introducing additional SA-based cost functions into the optimization. For example, one could use SA-based cost functions ensuring, for example,  $ST_{x,y,z}(D_{95\%,CTV}) < 5\%$  and  $ST_{R_{abs},R_{rel}}(D_{95\%,CTV}) < 5\%$  while allowing for larger sensitivity values in the biological inputs. By using the total effects also interactions between physical and RBE uncertainties are taken into account. Once the necessary performance for the execution of the variance-based SA during inverse planning is achieved, this will allow a systematic approach to physically and biologically robust IMPT planning. The consecutive step should then be followed by an evaluation of clinically relevant scenarios with focus on achievable improvements in proton therapy planning.

## 5. CONCLUSIONS

A framework for global, variance-based sensitivity analysis of proton therapy treatment plans has been implemented and demonstrated for two different variable RBE models. It is a powerful and flexible tool to assess the combined impact and interactions of positioning, range, and RBE uncertainties. Besides resulting overall uncertainties, the method provides quantitative information on the relative impact of the different input factors, which might have implications for future biologically robust IMPT planning.

## ACKNOWLEDGMENT

This project was supported by the DFG grant KA 4346/1-1 and the DFG Cluster of Excellence Munich Center for Advanced Photonics (MAP). Open access funding enabled and organized by ProjektDEAL.

## CONFLICT OF INTEREST

The authors have no conflict of interest to disclose.

## APPENDIX A

The following paragraphs describe how the individual sources of uncertainty are modeled.



## A. MODELING OF PATIENT SHIFTS

In the particle extension of CERR, the dose calculation is performed as follows:

$$D_{ij} = PDD_{E_j}(RD_{ij}) \cdot \frac{1}{2\pi \cdot \sigma_{E_j}^2(RD_{ij})} \cdot \exp\left(-\frac{r_{ij}^2}{2 \cdot \sigma_{E_j}^2(RD_{ij})}\right) \quad (\text{A1})$$

where  $PDD_{E_j}$  and  $\sigma_{E_j}^2$  are precalculated look-up tables for the depth-dose-curve and the lateral dose spread in water for incident proton energy  $E_j$  in water,  $RD_{ij}$  denotes the radiological depth on the central beam axis of the  $j$ -th spot at the depth of the  $i$ -th voxel, and  $r_{ij}$  is the distance of the  $i$ -th voxel to the central beam axis of the  $j$ -th spot. The lookup tables  $PDD_{E_j}$  and  $\sigma_{E_j}^2$  were precalculated in water using the Monte Carlo algorithm Geant4 for all relevant incident energies (50 to 260 MeV in steps of 1 MeV) assuming generic, mono-energetic beams.

Once the  $D_{ij}$  matrix is obtained, the dose in the  $i$ -th voxel can be calculated as:

$$d_i = \sum_j D_{ij} \omega_j \quad (\text{A2})$$

With a vector containing the pencil beam scanning (PBS) spot weights  $\omega$ . Therefore, the full dose vector can be obtained by a matrix-vector multiplication. In the particle extension of CERR, this is used for treatment plan optimization: Eq. (A2) is evaluated repeatedly to find the optimal weight vector. Throughout the entire optimization,  $D_{ij}$ , which contains all geometric information, is kept constant.

In order to recalculate treatment plans for a changed geometry, a fast way to generate the changed influence matrix is required. To achieve this, the following approximations are made: First, nondivergent beams are assumed, which is equivalent to assuming that the source is far away from the patient. The second approximation is that no patient deformations occur and the third that only rigid translations (isocenter shifts) are modeled, excluding rotations.

Then,  $r_{ij}$  and  $RD_{ij}$  in Eq. (A1) have to be replaced in order to reflect a patient shift. When  $r_{ij}$  is expressed in a two-dimensional Cartesian coordinate with axes perpendicular to the beam (i.e., a beams-eye-view (BEV) coordinate system):

$$r_{ij}^2 = x_{ij}^2 + y_{ij}^2 \quad (\text{A3})$$

The changes to the lateral offsets  $x_{ij}$  and  $y_{ij}$  caused by a rigid patient shift are then, in the approximation of nondivergent beams, the projection of the shift onto these axis:

$$x'_{ij} = x_{ij} + \Delta x(\varphi_n) \quad (\text{A4})$$

$$y'_{ij} = y_{ij} + \Delta y(\varphi_n) \quad (\text{A5})$$

where  $x'_{ij}$  and  $y'_{ij}$  denote the updated lateral offsets, which, for a given patient shift and in the approximation of nondivergent beams, only depend on the gantry angle  $\varphi_n$  of the beam  $n$ , making the computation very efficient for large  $ij$ -matrices,

since all entries belonging to the same beam can be treated in the same manner.

$RD_{ij}$  can be updated quickly by precalculating a set of neighboring raytracings of the original beam axis. Those raytracings are performed parallel to the original beam axis. The positions are defined on a 2D regular grid in the BEV coordinate system. During dose calculation, the algorithm has to select the correct neighbor raytracing  $RD_{ij}^*$  to replace  $RD_{ij}$  based on the lateral offsets  $\Delta x$  and  $\Delta y$  from Eqs. (A4) and (A5). This method is similar to the “virtual beamlets” approach proposed by Unkelbach et al.,<sup>38</sup> with the difference, that in our approach only the raytracing is approximated by the neighbor which is closest to the patient shift, while the lateral part in Eq. (A1) is modeled exactly according to Eqs. (A4) and (A5). The approach to keep precalculated neighboring raytracings for every PBS spot is memory intensive, however, since only small setup errors are expected and need to be treated in the SA, sets of only a few shifts of up to a few millimeters are sufficient. In our calculations, we used a 5 by 5 grid of raytracings (where the raytracing at the center corresponded to the nominal case without setup error) with a grid constant of 2 mm, therefore setup errors of up to about 4 mm can be handled. Since in our patient cases described below we assumed the setup error to be normal distributed in all three spatial dimensions with standard deviation 1 mm, this is sufficient for our purposes. If necessary, additional grid points can be included to support larger shifts at the cost of an increased memory usage.

## B. MODELING OF RANGE UNCERTAINTY

Range uncertainty is modeled by applying a further transformation to the radiological depth:

$$RD_{ij} = (1 + \Delta RD_{rel}) \cdot RD_{ij}^* + \Delta RD_{abs} \quad (\text{A6})$$

where  $RD_{ij}^*$  is the “neighbor raytracing” from the last paragraph,  $\Delta RD_{rel}$  is the relative error mostly associated with imperfect CT number to stopping power conversion and  $\Delta RD_{abs}$  is an absolute offset suited to model patient changes in the beam path (e.g., weight loss, filling of the paranasal sinus with liquid etc.).<sup>1,39</sup> In this approximation, the absolute range shift applied in one RWD calculation is the same for all PBS spots in one beam, therefore, perturbations only affecting parts of the beam cannot be modeled. The relative range shift is the same for all beams. This coupling between different beam directions is justified by the fact that the relative range shift is assumed to originate primarily in imperfect CT to stopping power conversion.

## C. MODELING OF RBE UNCERTAINTY

Currently, two models for variable RBE predictions are supported: the mechanistic RMF and the phenomenological Wedenberg model.

### C.1. RMF model

The repair-misrepair-fixation (RMF) model was introduced by Carlson et al.<sup>29</sup> In our implementation, it uses estimates from a Monte Carlo damage simulation (MCDS)<sup>31</sup> to link  $\alpha_P$  and  $\beta_P$  to double strand break (DSB) yields.<sup>30</sup> For a given particle type and energy, these are calculated as:

$$\alpha_P = \frac{\Sigma}{\Sigma_X} \left[ \alpha_X + 2 \frac{\beta_X}{\Sigma_X} (\Sigma \cdot \bar{z}_F - \Sigma_X \cdot \bar{z}_{F,X}) \right] \quad (\text{A7})$$

$$\sqrt{\beta_P} = \sqrt{\beta_X} \cdot \frac{\Sigma}{\Sigma_X} \quad (\text{A8})$$

where  $\alpha_X$  and  $\beta_X$  are the reference x-ray reference radiosensitivity parameters,  $\Sigma$  and  $\Sigma_X$  are the DSB yields for the particle (defined as the initial number of DSB per Gray per giga base pair  $\text{Gy}^{-1} \text{Gbp}^{-1}$ ) and reference radiation of a Co-60 source, respectively; and  $\bar{z}_F$  and  $\bar{z}_{F,X}$  denote the frequency-mean specific energy. The default RMF model settings were used for the predictions. The model does not require any fit to experimental data. The MCDS software version 3.10A<sup>31</sup> with default settings (cell nucleus diameter 5  $\mu\text{m}$ ) was used to calculate DSB yields and the frequency-mean specific energy for all relevant proton energies. The DSB yields and frequency-mean specific energy were used as inputs for the RMF model as previously demonstrated by Carlson et al.,<sup>29</sup> Frese et al.<sup>30</sup> and Kamp et al.<sup>25</sup>

A rapid implementation<sup>25,26</sup> of this model in the  $ij$ -formalism for carbon ions has already been used for a variance-based SA of biological uncertainties by Kamp et al.<sup>15</sup> and for RWD optimization by Guan et al.<sup>40</sup>. We use a similar approach. To obtain  $\alpha_{P,ij}$  and  $\beta_{P,ij}$  (the radiosensitivity parameters for the dose contribution of the  $j$ -th spot to the  $i$ -th voxel), Eqs. (A7) and (A8) need to be integrated against the fluence spectrum  $\phi_{ij}$  and the stopping power  $SP$ , which are both a function of the particle energy  $E$ . By simulating the fluence spectra using Geant4 as described above, precalculated, tabulated data for the resulting integrals can be obtained.

Integration of Eqs. (A7) and (A8) against the fluence spectrum  $\Phi_{ij}$  and the stopping power  $SP(E)$  yields:

$$\alpha_{P,ij} = \alpha_{X,i} \frac{\int_0^\infty \frac{\Sigma(E)}{\Sigma_X} \cdot \Phi_{ij}(E) \cdot SP(E) dE}{\int_0^\infty \Phi_{ij}(E) \cdot SP(E) dE} + \frac{\int_0^\infty 2\beta_{X,i} \cdot \left[ \frac{\Sigma(E)}{\Sigma_X} \right]^2 \cdot \bar{z}_F(E) \cdot \Phi_{ij}(E) \cdot SP(E) dE}{\int_0^\infty \Phi_{ij}(E) \cdot SP(E) dE} \quad (\text{A9})$$

$$\sqrt{\beta_{P,ij}} = \sqrt{\beta_{X,i}} \cdot \frac{\int_0^\infty \frac{\Sigma(E)}{\Sigma_X} \cdot \Phi_{ij}(E) \cdot SP(E) dE}{\int_0^\infty \Phi_{ij}(E) \cdot SP(E) dE} \quad (\text{A10})$$

By introducing precalculated constants for the integrals:

$$\frac{\int_0^\infty \frac{\Sigma(E)}{\Sigma_X} \cdot \Phi_{ij}(E) \cdot SP(E) dE}{\int_0^\infty \Phi_{ij}(E) \cdot SP(E) dE} = C_{1,ij} \quad (\text{A11})$$

$$\frac{\int_0^\infty 2 \cdot \left[ \frac{\Sigma(E)}{\Sigma_X} \right]^2 \cdot \bar{z}_F(E) \cdot \Phi_{ij}(E) \cdot SP(E) dE}{\int_0^\infty \Phi_{ij}(E) \cdot SP(E) dE} = C_{2,ij}^* \quad (\text{A12})$$

equations (A9) and (A10) can be written as:

$$\begin{aligned} \alpha_{P,ij} &= \alpha_{X,i} \cdot C_{1,ij} + \beta_{X,i} \cdot C_{2,ij}^* - 2\beta_{X,i} \cdot \bar{z}_{F,X} \cdot C_{1,ij} \\ &= (\alpha_{X,i} - 2\beta_{X,i} \cdot \bar{z}_{F,X}) \cdot C_{1,ij} + \beta_{X,i} \cdot C_{2,ij}^* \end{aligned} \quad (\text{A13})$$

$$\sqrt{\beta_{P,ij}} = \sqrt{\beta_{X,i}} \cdot C_{1,ij} \quad (\text{A14})$$

This reparametrization is very closely related to the one used by Kamp et al.,<sup>26</sup> who used the two reparametrization constants  $C_{1,ij}$  and  $C_{2,ij}$ . While in our case  $C_{1,ij}$  is defined in exactly the same way, note that our  $C_{2,ij}^*$  is related to  $C_{2,ij}$  via:

$$C_{2,ij} = C_{2,ij}^* - 2 \cdot \bar{z}_{F,X} \cdot C_{1,ij} \quad (\text{A15})$$

The precalculated tables  $C_1$  and  $C_2^*$  are referenced with the radiological depth and the corresponding incident proton energy to obtain  $C_{1,ij}$  and  $C_{2,ij}^*$ :

$$C_{1,ij} = C_{1,E_j}(RD_{ij}) \quad (\text{A16})$$

$$C_{2,ij}^* = C_{2,E_j}^*(RD_{ij}) \quad (\text{A17})$$

To model biological uncertainty in the RMF model, we will treat  $\Sigma$  as uncertain and apply a relative variation  $\left(\frac{\Delta\Sigma}{\Sigma}\right)$ . Under the assumption, that this variation is independent of the energy,  $C_{1,ij}$  depends linearly on  $\left(\frac{\Delta\Sigma}{\Sigma}\right)$ . For  $C_{2,ij}^*$ , there is a quadratic dependence. Therefore, Eqs. (A13) and (A14) become:

$$\begin{aligned} \alpha'_{P,ij} \left( \frac{\Delta\Sigma}{\Sigma} \right) &= (\alpha_{X,i} - 2\beta_{X,i} \bar{z}_{F,X}) \cdot C_{1,ij} \cdot \left( 1 + \frac{\Delta\Sigma}{\Sigma} \right) + \\ &\quad \beta_{X,i} \cdot C_{2,ij}^* \cdot \left( 1 + \frac{\Delta\Sigma}{\Sigma} \right)^2 \end{aligned} \quad (\text{A18})$$

$$\sqrt{\beta'_{P,ij} \left( \frac{\Delta\Sigma}{\Sigma} \right)} = \sqrt{\beta_{X,i}} \cdot C_{1,ij} \cdot \left( 1 + \frac{\Delta\Sigma}{\Sigma} \right) \quad (\text{A19})$$

### C.2. Wedenberg model

In the Wedenberg model,<sup>32</sup>  $\alpha_P$  is assumed to depend on the x-ray reference radiosensitivity parameter and increase linearly with the linear energy transfer (LET)::

$$\frac{\alpha_{P,ij}}{\alpha_{X,i}} = 1 + \frac{q \cdot LET_{ij}}{(\alpha/\beta)_{X,i}} \quad (\text{A20})$$

where  $q = 0.434 \text{ Gy} \frac{\mu\text{m}}{\text{keV}}$  is obtained from a fit to in-vitro cell survival data and  $LET_{ij}$  is the LET contribution of the  $j$ -th PBS spot to the  $i$ -th voxel. It is obtained by referencing precalculated depth-LET tables for the incident proton energies  $E_j$  with the radiological depth  $RD_{ij}$ :

$$LET_{ij} = LET_{E_j}(RD_{ij}) \quad (A21)$$

where  $LET_{E_j}$  contains MC calculated data for the dose-weighted LET in water simulated using Geant4 as described above.  $\beta_P$  is assumed to be equal to the reference value:

$$\frac{\beta_{P,ij}}{\beta_{X,i}} = 1 \quad (A22)$$

To include uncertainty in this model, both  $q$  and  $\beta_{P,ij}$  were treated as uncertain:

$$q' = q \cdot \left(1 + \frac{\Delta q}{q}\right) \quad (A23)$$

$$\beta'_{P,ij} = \beta_{P,ij} \cdot \left(1 + \frac{\Delta \beta_P}{\beta_{P,ij}}\right) \quad (A24)$$

For both RBE models,  $\alpha_{P,i}$  and  $\sqrt{\beta_{P,i}}$  for the  $i$ -th voxel can then be calculated as dose-weighted sums in the  $ij$ -formalism<sup>41</sup>:

$$\alpha_{P,i} = \frac{1}{d_i} \sum_j \alpha_{P,ij} \cdot D_{ij} \cdot \omega_j \quad (A25)$$

$$\sqrt{\beta_{P,i}} = \frac{1}{d_i} \sum_j \sqrt{\beta_{P,ij}} \cdot D_{ij} \cdot \omega_j \quad (A26)$$

Then, the RBE is calculated using the formula:

$$RBE_i(\alpha_{X,i}, \beta_{X,i}, \alpha_{P,i}, \beta_{P,i}, d) = \frac{-\alpha_{X,i} + \sqrt{\alpha_{X,i}^2 + 4\beta_{X,i}d_i \cdot (\alpha_{P,i} + \beta_{P,i}d_i)}}{2\beta_{X,i}d_i} \quad (A27)$$

$RBE_i(\alpha_{X,i}, \beta_{X,i}, \alpha_{P,i}, \beta_{P,i}, d)$  reduces to  $RBE_i(\alpha_{X,i}/\beta_{X,i}, LET_i, q, d)$  for the Wedenberg model and to  $RBE_i(\alpha_{X,i}/\beta_{X,i}, C_{1,i}, C_{2,i}^*, d)$  for the RMF model, therefore the uncertainty in the x-ray reference parameters can be treated by varying one parameter, the fraction  $\alpha_{X,i}/\beta_{X,i}$ .

A convenient property of the  $ij$ -formalism is that the RWD calculation can be restricted to arbitrary subgroups regions of interest to reduce memory usage. In our calculations we typically restricted the RWD calculation to a 2 cm expansion of the CTV and all OARs for whom the DVH quantiles were included in the SA (optic nerves, chiasm, brain stem).

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: jan.hofmaier@med.uni-muenchen.de; Telephone: 004989440076744.

## REFERENCES

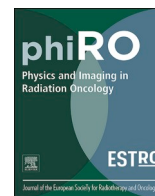
- Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Phys Med Biol.* 2012;57:R99–R117.
- Schuemann J, Dowdell S, Grassberger C, Min CH, Paganetti H. Site-specific range uncertainties caused by dose calculation algorithms for proton therapy. *Phys Med Biol.* 2014;59:4007–4031.
- Lomax AJ. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties. *Phys Med Biol.* 2008;53:1027–1042.
- 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;53:1043–1056.
- Paganetti H. Relative biological effectiveness (RBE) values for proton beam therapy. Variations as a function of biological endpoint, dose, and linear energy transfer. *Phys Med Biol.* 2014;59:R419–R472.
- Mohan R, Peeler CR, Guan F, Bronk L, Cao W, Grosshans DR. Radiobiological issues in proton therapy. *Acta Oncol (Madr).* 2017;56:1367–1373.
- Paganetti H. Proton relative biological effectiveness-uncertainties and opportunities. *Int J Part Ther.* 2018;5:2–14.
- Bangert M, Hennig P, Oelfke U. Analytical probabilistic modeling for radiation therapy treatment planning. *Phys Med Biol.* 2013;58:5401–5419.
- Perkó Z, van der Voort SR, van de Water S, Hartman CMH, Hoogeman M, Lathouwers D. Fast and accurate sensitivity analysis of IMPT treatment plans using Polynomial Chaos Expansion. *Phys Med Biol.* 2016;61:4646–4664.
- Unkelbach J, Alber M, Bangert M, et al. Robust radiotherapy planning. *Phys Med Biol.* 2018;63:22TR02.
- Unkelbach J, Paganetti H. Robust proton treatment planning: physical and biological optimization. *Semin Radiat Oncol.* 2018;28:88–96.
- Sobol IM. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math Comput Simul.* 2001;55:271–280.
- Borgonovo E, Plischke E. Sensitivity analysis: a review of recent advances. *Eur J Oper Res.* 2016;248:869–887.
- Kamp F, Brüningk S, Cabal G, Mairani A, Parodi K, Wilkens JJ. Variance-based sensitivity analysis of biological uncertainties in carbon ion therapy. *Phys Medica.* 2014;30:583–587.
- Kamp F, Wilkens JJ. Application of variance-based uncertainty and sensitivity analysis to biological modeling in carbon ion treatment plans. *Med Phys.* 2019;46:437–447.
- Zvereva A, Kamp F, Schlattl H, Zankl M, Parodi K. Impact of interpatient variability on organ dose estimates according to MIRD schema: uncertainty and variance-based sensitivity analysis. *Med Phys.* 2018;45:3391–3403.
- Sobol IM. Sensitivity estimates for nonlinear mathematical models. *Math Model Comput Exp.* 1993;1:407–414.
- Homma T, Saltelli A. Importance measures in global sensitivity analysis of nonlinear models. *Reliab Eng Syst Saf.* 1996;52:1–17.
- Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput Phys Commun.* 2010;181:259–270.
- Sobol IM. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput Math Math Phys.* 1967;7:86–112.
- Sobol' I, Turchaninov V, Levitan Y, Shukhman B. Quasi-random sequence generators. *Ipm zak;* 1992:30.
- Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research. *Med Phys.* 2003;30:979–985.
- Schell S, Wilkens JJ. Advanced treatment planning methods for efficient radiation therapy with laser accelerated proton and ion beams. *Med Phys.* 2010;37:5330–5340.
- Brüningk SC, Kamp F, Wilkens JJ. EUD-based biological optimization for carbon ion therapy. *Med Phys.* 2015;42:6248–6257.
- Kamp F, Cabal G, Mairani A, Parodi K, Wilkens JJ, Carlson DJ. Fast biological modeling for voxel-based heavy ion treatment planning using the mechanistic repair-misrepair-fixation model and nuclear fragment spectra. *Int J Radiat Oncol Biol Phys.* 2015;93:557–568.
- Kamp F, Carlson DJ, Wilkens JJ. Rapid implementation of the repair-misrepair-fixation (RMF) model facilitating online adaption of radiosensitivity parameters in ion therapy. *Phys Med Biol.* 2017;62:N285–N296.
- Lühr A, von Neubeck C, Pawelke J, et al. “Radiobiology of proton therapy”: results of an international expert workshop. *Radiother Oncol.* 2018;128:56–67.
- Paganetti H, Blakely E, Carabe-Fernandez A, et al. Report of the AAPM TG-256 on the relative biological effectiveness of proton beams in radiation therapy. *Med Phys.* 2019;46:e53–e78.
- Carlson DJ, Stewart RD, Semenenko VA, Sandison GA. Combined use of Monte Carlo DNA damage simulations and deterministic repair

- models to examine putative mechanisms of cell killing. *Radiat Res.* 2008;169:447–459.
30. Frese MC, Yu VK, Stewart RD, Carlson DJ. A mechanism-based approach to predict the relative biological effectiveness of protons and carbon ions in radiation therapy. *Int J Radiat Oncol Biol Phys.* 2012;83:442–450.
  31. Stewart RD, Yu VK, Georgakilas AG, Koumenis C, Park JH, Carlson DJ. Effects of radiation quality and oxygen on clustered DNA lesions and cell death. *Radiat Res.* 2011;176:587–602.
  32. Wedenberg M, Lind BK, Hårdemark B. A model for the relative biological effectiveness of protons: the tissue specific parameter  $\alpha/\beta$  of photons is a predictor for the sensitivity to LET changes. *Acta Oncol (Madr).* 2013;52:580–588.
  33. Mesías MC, Boda-Heggemann J, Thoelking J, Lohr F, Wenz F, Wertz H. Quantification and assessment of interfraction setup errors based on cone beam CT and determination of safety margins for radiotherapy. *PLoS One.* 2016;11:e0150326.
  34. Wohlfahrt P, Möhler C, Stützer K, Greilich S, Richter C. Dual-energy CT based proton range prediction in head and pelvic tumor patients. *Radiother Oncol.* 2017;125:526–533.
  35. Schulz-Ertner D, Karger CP, Feuerhake A, et al. Effectiveness of carbon ion radiotherapy in the treatment of skull-base chordomas. *Int J Radiat Oncol Biol Phys.* 2007;68:449–457.
  36. Wieser HP, Hennig P, Wahl N, Bangert M. Analytical probabilistic modeling of RBE-weighted dose for ion therapy. *Phys Med Biol.* 2017;62:8959–8982.
  37. Semenenko VA, Stewart RD. A fast Monte Carlo algorithm to simulate the spectrum of DNA damages formed by ionizing radiation. *Radiat Res.* 2004;161:451–457.
  38. 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;36:149–163.
  39. Taasti VT, Bäumer C, Dahlgren CV, et al. Inter-centre variability of CT-based stopping-power prediction in particle therapy: survey-based evaluation. *Phys Imaging Radiat Oncol.* 2018;6:25–30.
  40. Guan F, Geng C, Carlson DJ, et al. A mechanistic relative biological effectiveness model-based biological dose optimization for charged particle radiobiology studies. *Phys Med Biol.* 2018;64:015008.
  41. Wilkens JJ, Oelfke U. Fast multifield optimization of the biological effect in ion therapy. *Phys Med Biol.* 2006;51:3127–3140.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Data S1.** Application of the SA framework to an additional prostate case.



## Technical Note

# Combining inter-observer variability, range and setup uncertainty in a variance-based sensitivity analysis for proton therapy

Jan Hofmaier<sup>a,\*</sup>, Franziska Walter<sup>a</sup>, Indrawati Hadi<sup>a</sup>, Maya Rottler<sup>a</sup>, Rieke von Bestenbostel<sup>a</sup>, George Dedes<sup>c</sup>, Katia Parodi<sup>c</sup>, Maximilian Niyazi<sup>a</sup>, Claus Belka<sup>a,d</sup>, Florian Kamp<sup>a,b</sup>

<sup>a</sup> Department of Radiation Oncology, University Hospital, LMU Munich, Munich, Germany

<sup>b</sup> Department of Radiation Oncology and CyberKnife Center, Faculty of Medicine, University Hospital Cologne, Cologne, Germany

<sup>c</sup> Department of Medical Physics, Faculty of Physics, LMU Munich, Munich, Germany

<sup>d</sup> German Cancer Consortium (DKTK), Munich, Germany



## ARTICLE INFO

## Keywords:

Proton therapy  
Range uncertainty  
Setup uncertainty  
Inter-observer variability  
Sensitivity analysis

## ABSTRACT

Margin concepts in proton therapy aim to ensure full dose coverage of the clinical target volume (CTV) in presence of setup and range uncertainty. Due to inter-observer variability (IOV), the CTV itself is uncertain. We present a framework to evaluate the combined impact of IOV, setup and range uncertainty in a variance-based sensitivity analysis (SA). For ten patients with skull base meningioma, the mean calculation time to perform the SA including  $1.6 \times 10^4$  dose recalculations was 59 min. For two patients in this dataset, IOV had a relevant impact on the estimated CTV  $D_{95\%}$  uncertainty.

## 1. Introduction

Treatment plans in proton therapy are affected by range and setup uncertainties. These are typically compensated through margin concepts or robust planning approaches. Margin concepts aim at covering the clinical target volume (CTV) in presence of range and setup uncertainty [1]. However, due to inter-observer variability (IOV), the CTV itself is uncertain. While there are many studies assessing IOV, only few studies have investigated dosimetric consequences of IOV [2], e.g. Lobefalo et al. [3] who investigated the dosimetric impact of IOV in three-dimensional conformal radiotherapy (3D-CRT) and volumetric modulated arc therapy for rectal tumours, Hellebust et al. [4] who assessed the dosimetric impact of IOV in brachytherapy for cervical cancer and Eminowicz et al. [5], who studied the dosimetric impact of IOV in VMAT for cervical cancer. To the best of our knowledge, there is no study assessing the combined and relative impact of range, setup uncertainty and IOV in proton therapy in a quantitative way. The statistical method of variance-based sensitivity analysis (SA) is suited for this, since it can be used to assess the impact of uncertainty of multiple input parameters on the output of a quantitative model [6]. In the context of patient dose calculation in medical physics, the technique has been previously applied to relative biological effectiveness (RBE) uncertainties in carbon ion therapy [7,8] and to estimate the impact of interpatient variability

on organ dose estimates in nuclear medicine [9]. Recently, a framework to evaluate the combined impact of range, setup and RBE uncertainty in a variance-based SA has been presented by our group [10]. In this technical note, an extension of the framework to include IOV is shown. The feasibility of the approach was demonstrated by using it to investigate the relative impact of IOV, range and setup uncertainty on proton plans for a dataset with ten patients with skull base meningioma.

## 2. Materials and methods

## 2.1. Variance-based sensitivity analysis

In the Monte Carlo method of global variance-based SA, the output of a model  $Y = f(X)$  with  $k$  input factors  $X = (x_1, x_2, \dots, x_k)$  which are subject to uncertainty is recalculated many times while simultaneously and randomly varying the input factors within their assumed distributions. In our particular case, the model  $f(X)$  corresponded to a dose calculation followed by a dose volume histogram (DVH) calculation. The output  $Y$  corresponded to DVH parameters of interest. The input factors  $(x_1, x_2, \dots, x_k)$  included patient shifts in three spatial dimensions, absolute and relative range shifts as well as IOV, resulting in  $k = 6$  input factors. The resulting variance  $V(Y)$  is decomposed as [6]:

\* Corresponding author.

E-mail address: [jan.hofmaier@med.uni-muenchen.de](mailto:jan.hofmaier@med.uni-muenchen.de) (J. Hofmaier).

<https://doi.org/10.1016/j.phro.2021.11.005>

Received 30 September 2021; Received in revised form 16 November 2021; Accepted 16 November 2021

Available online 2 December 2021

2405-6316/© 2021 The Author(s). Published by Elsevier B.V. on behalf of European Society of Radiotherapy & Oncology. This is an open access article under the

CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



$$V(Y) = \sum_{l=1}^k V_l + \sum_{l=1}^k \sum_{m>l}^k V_{lm} + \sum_{l=1}^k \sum_{m>l}^k \sum_{n>m}^k V_{lmn} + \dots + V_{1\dots k} \quad (1)$$

resulting in  $(2^k - 1)$  terms. The first order terms are

$$V_l = V[E(Y|X_l)] \quad (2)$$

The expectation value  $E(Y|X_l)$  is hereby calculated over all possible values of all input factors except for  $X_l$ , which is kept fixed. The second order terms, which are representing the interaction between the inputs  $X_l$  and  $X_m$ , are

$$V_{lm} = V[E(Y|X_l, X_m)] - V_l - V_m \quad (3)$$

Higher order terms are defined in an analogous fashion. Sensitivity indices are defined by normalising to the overall variance

$$S_l = \frac{V_l}{V(Y)} \quad (4)$$

$$S_{lm} = \frac{V_{lm}}{V(Y)} \quad (5)$$

and so on. Total effect indices are defined by summing all terms of any order containing  $l$ :

$$ST_l = S_l + \sum_{m \neq l} S_{lm} + \dots + S_{1\dots k} \quad (6)$$

Like in a previous study from our group [10], the efficient Monte Carlo method proposed by Saltelli [6] was used for direct calculation of  $S_l$  and  $ST_l$ , and sampling from low-discrepancy quasi-random sequences was employed to improve convergence. This method requires  $N(k+2)$  model evaluations, where  $N$  is typically of the order of  $10^3$ . In our study, as described above, we had  $k = 6$  input factors. We set  $N = 2048$ , which resulted in approximately  $1.6 \cdot 10^4$  model evaluations. The sensitivity analysis framework was extended to include IOV. Additionally to the fast, graphics processing unit (GPU) based pencil beam algorithm capable of modeling setup and range variations described in the previous publication from our group [10], the possibility to include multiple treatment plans and to switch randomly between them was added.

## 2.2. Clinical dataset

Datasets of ten patients with benign (WHO grade I) meningioma of the skull base were included in this study. For all patients, contrast enhanced magnetic resonance imaging (MRI) and DOTATATE positron emission tomography (PET) images were available in addition to a planning computed tomography (CT).

## 2.3. Target delineation and treatment planning

A rigid image registration of MRI, PET and planning CT images was performed. For each patient, four clinicians independently delineated the gross tumor volume (GTV) taking into account all imaging modalities (GTV<sub>observer</sub>). A consensus GTV (GTV<sub>STAPLE</sub>) was created using the simultaneous truth and performance level estimation (STAPLE) algorithm [11] in the research treatment planning system computational environment for radiological research (CERR) [12]. This implementation of an expectation-maximization algorithm generates a probabilistic estimate of the true volume based on the volumes delineated by multiple observers. The GTV<sub>STAPLE</sub> was used as the "ground truth" GTV. As an example, the four GTV<sub>observer</sub> and the GTV<sub>STAPLE</sub> contours for patient number 1 are shown in the supplementary material. The CTV<sub>observer</sub> and the CTV<sub>STAPLE</sub> were defined as the respective GTV without any margins applied (i.e. GTV = CTV), as suggested in a current guideline [13]. To obtain the planning target volumes (PTVs), gantry-angle specific margins were applied. To compensate for proton range uncertainty, larger

margins were applied in beam direction than laterally. The applied margins were 6, 5 and 3 mm in distal, proximal and lateral directions, respectively. For a typical margin recipe of 3.5% + 3 mm, the distal margin of 6 mm would correspond to a radiological target depth of approximately 9 cm. Since all tumours were at the skull base and therefore at similar depths, the same absolute margins were applied to all patients for simplicity. For each CTV<sub>observer</sub> a PTV<sub>observer</sub> was created. For each PTV<sub>observer</sub> of each patient a spot scanning proton treatment plan with one beam was generated using non-robust optimization, resulting in a total number of 40 treatment plans (four treatment plans for each of the ten patients). The gantry angle was chosen individually for each patient. The proton plans were optimized to deliver 1.8 Gy (RBE) per fraction to the PTV<sub>observer</sub>. A spatially constant RBE of 1.1 was assumed.

## 2.4. Application of the SA framework

Like in the previous study from our group [10], the variance-based SA was performed assuming the following uncertainty distributions for the input factors mentioned in Section 2.1: For patient shifts in X, Y and Z directions, a normal distribution with standard deviation  $\sigma_{X,Y,Z} = 1$  mm truncated to  $2\sigma_{X,Y,Z}$  was assumed. For relative range shifts the probability density was set to a normal distribution with standard deviation  $\sigma_{r,rel} = 3\%$  truncated to  $2\sigma_{r,rel}$ . Additionally, absolute range shifts following a normal distribution with standard deviation  $\sigma_{r,abs} = 1$  mm truncated to  $2\sigma_{r,abs}$  were assumed. For IOV, an equal probability of  $p = 0.25$  for each of the four observer treatment plans was chosen. To perform the SA, the dose distribution was re-calculated approximately  $1.6 \cdot 10^4$  times (corresponding to  $N = 2048$  and  $k = 6$  in the Saltelli formalism, as described in Section 2.1) while simultaneously sampling from the above uncertainty distributions. An Nvidia Quadro RTX 8000 GPU with 48 gigabytes of memory was used. For the resulting dose distributions, DVHs were calculated for the CTV<sub>STAPLE</sub>. Confidence intervals (CIs) and sensitivity indices for the dose level enclosing 95% of the CTV<sub>STAPLE</sub> ( $D_{95\%}$ ) were calculated. Convergence plots of the sensitivity indices were created. The obtained total effect indices  $ST$  were converted to  $SI_{IOV}$ , the sum of all interaction terms with involvement of IOV and  $SI_{other}$ , the sum of all interaction terms without involvement of IOV. By definition is

$$SI_{IOV} = ST_{IOV} - S_{IOV} \quad (7)$$

and due to normalization

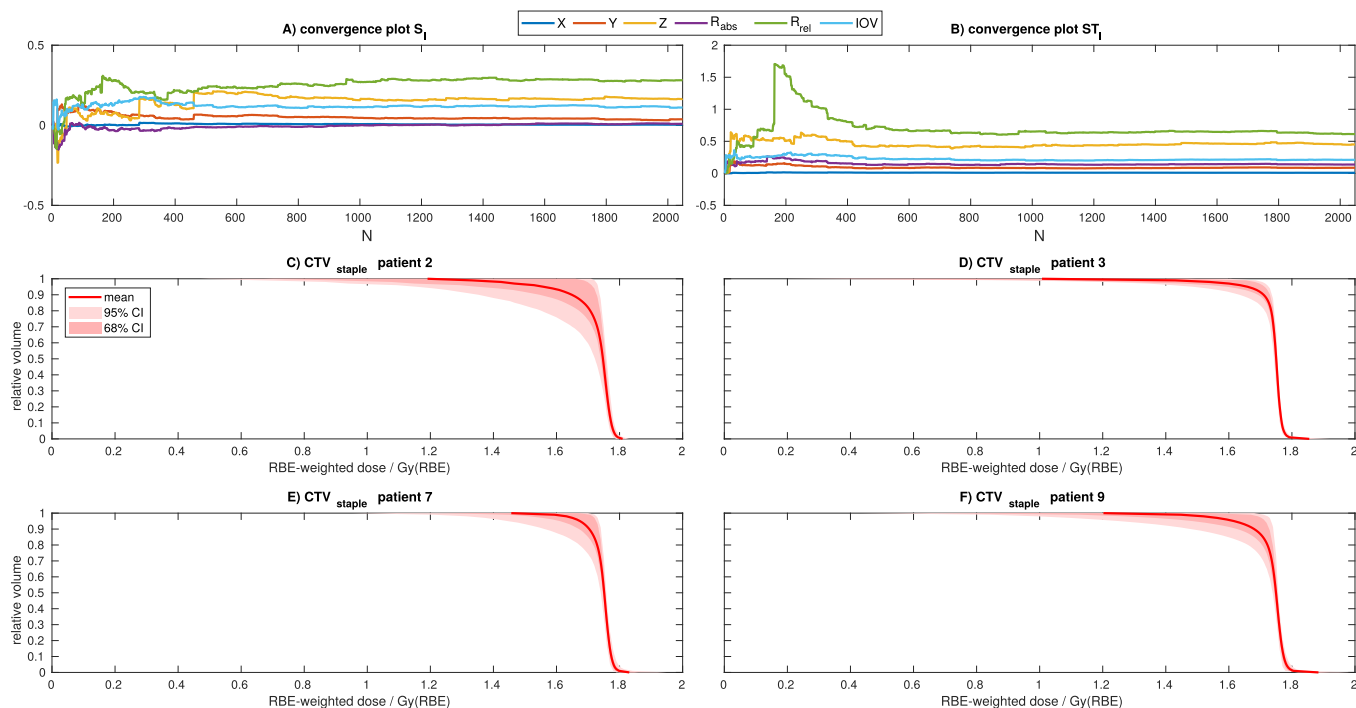
$$SI_{other} = 1 - S_{setup} - S_{range} - ST_{IOV} \quad (8)$$

## 3. Results

The mean calculation time to perform the  $1.6 \cdot 10^4$  dose calculations was 59 min. Large differences were observed for the calculation times for different patients, which ranged from 11 min to 195 min. Convergence plots for  $S_l$  and  $ST_l$  for an exemplary patient are shown in panels A and B of Fig. 1. By visual inspection of the convergence plots it becomes evident that a sufficient convergence was achieved well below  $N = 2048$ .

Results for the  $D_{95\%}$  are presented in Table 1. For six patients, the width of the  $CI_{95\%}$  for the  $D_{95\%}$  was below 0.18 Gy (10% of the prescribed dose of 1.8 Gy). Uncertainties of more than 10 % were observed for patients 2, 3, 7 and 9. Here the widths of the  $CI_{95\%}$  for the  $D_{95\%}$  were 0.57, 0.24, 0.28 and 0.48 Gy, respectively. Plots of the DVHs for the CTV<sub>STAPLE</sub> for these four patients with their corresponding 95 % and 68 % CIs are shown in panels C to F of Fig. 1. For two of these patients, the overall influence of IOV was negligible ( $SI_{IOV} + SI_{other} < 0.05$  for patients 7 and 9). In both cases, range uncertainty was the most important contribution to overall uncertainty ( $S_{range}$  was 0.53 and 0.70 for patients 7 and 9, respectively). For patients 2 and 3, however, IOV played a major





**Fig. 1.** Convergence plots for  $S_1$  and  $ST_1$  for one patient (panels A and B) and DVHs for  $CTV_{STAPLE}$  for the four patients with the largest overall  $D_{95\%}$  uncertainties (panels C to F). The variability of the DVH in presence of setup uncertainty, range uncertainty and IOV is visualized by the shaded areas (68% and 95% CIs). The solid line indicates the mean value over all simulated error scenarios.

**Table 1**

Uncertainty and sensitivity analysis results for the  $D_{95\%}$  for  $(CTV)_{STAPLE}$ . For each patient, the mean value and 95% and the 68% CIs have been calculated. The relative contribution to the overall uncertainty is broken down to first order indices  $S_{setup}$ ,  $S_{range}$  and  $S_{IOV}$ , higher order indices with involvement of IOV ( $SI_{IOV}$ ) and higher order indices without involvement of IOV ( $SI_{other}$ ).

pat.	mean [Gy]	$CI_{95\%}$ [Gy]	$CI_{68\%}$ [Gy]	$S_{setup}$	$S_{range}$	$S_{IOV}$	$SI_{IOV}$	$SI_{other}$
1	1.71	1.62–1.74	1.69–1.73	0.21	0.29	0.11	0.10	0.29
2	1.56	1.16–1.73	1.35–1.72	0.12	0.35	0.34	0.09	0.10
3	1.66	1.49–1.73	1.59–1.71	0.14	0.11	0.62	0.01	0.12
4	1.70	1.59–1.73	1.68–1.72	0.19	0.14	0.10	0.12	0.45
5	1.73	1.71–1.74	1.72–1.74	0.37	0.31	0.00	0.14	0.18
6	1.73	1.70–1.74	1.72–1.73	0.20	0.29	0.03	0.04	0.44
7	1.67	1.45–1.73	1.61–1.73	0.25	0.53	0.00	0.02	0.20
8	1.69	1.58–1.74	1.65–1.73	0.23	0.49	0.04	0.05	0.19
9	1.62	1.25–1.73	1.48–1.71	0.13	0.70	0.01	0.00	0.16
10	1.73	1.70–1.74	1.72–1.74	0.26	0.24	0.02	0.06	0.42

role for overall uncertainty ( $S_{IOV} + SI_{IOV}$  was 0.43 and 0.63 for patients 2 and 3, respectively).

**4. Discussion**

A framework for the variance-based SA of setup, range and IOV has been presented. To the best of our knowledge, this study is the first to assess the relative dosimetric impact of setup uncertainty, range uncertainty and IOV in a variance-based SA. In a first analysis of ten patients, calculation times were of the order of a few minutes to a few hours. These calculation times are fast enough for offline plan evaluation. Although this was not investigated in this study, it can be assumed that the differences in calculation time were caused by differences in the sizes and depths of the target volumes. The convergence plots in Fig. 1 suggest that actually less than  $N = 2048$  would have been sufficient to achieve convergence, therefore the calculation times could be reduced by stopping the calculation after reaching a predefined convergence criterion. While for the majority of patients, the overall uncertainties in CTV coverage were small, in some cases the coverage was deteriorated. The dominating contributions to overall uncertainty were either range

uncertainty or IOV. This suggests that IOV might have a relevant effect on target coverage in some patients.

In this work, the analysis was restricted to skull base meningioma, since the framework does not support organ motion at the moment. Furthermore, a pencil beam algorithm was used, whose accuracy is known to decrease in regions of high heterogeneity. The framework would be applicable without modification to other tumour sites for which these limitations are acceptable. The possibility to model motion could be included by extending the framework to use multiple CT geometries (e.g. phases of a 4D-CT to model breathing motion), at the cost of an increased memory usage and longer calculation times. In the previous publication from our group [10], uncertainties in variable RBE models were evaluated in combination with setup and range uncertainty. In this study, since the focus was on IOV, RBE uncertainty was not taken into account and a constant RBE of 1.1 was assumed. However, the combined evaluation of all four types of uncertainty could in principle also be included in the framework. This could be used in future studies to assess the combined impact of range, setup and RBE uncertainty and IOV. The evaluation of the CTV  $D_{95\%}$  in presence of IOV required a "ground truth" CTV. Unfortunately, this volume is not known. In this

work, the consensus target volume created with the STAPLE algorithm was used to define a "ground truth" target volume, as has been done previously [14]. Since this algorithm provides a maximum likelihood estimate for the actual CTV based on the observer CTVs themselves, this approach is well suited to capture the variability within a group of observers. However, it cannot correct systematic deviations from the ground truth CTV within the observer group. Furthermore, in this study only data from four observers was available, which was considered sufficient to show the feasibility of the approach. However, outlier contours could have considerable effect on the evaluation. For this reason, both the number of patients and the number of observers needs to be increased for future systematic evaluations of the impact of IOV in combination with setup and range uncertainties. Another limitation is that in our study simple proton plans with only one beam direction were used. More clinically realistic plans with multiple beam directions are supported by the framework without modifications, but have higher memory requirements and will lead to longer calculation times.

In this technical note, no metrics of contour similarity such as Dice coefficients or Hausdorff distances were evaluated. The presented framework might be used in future studies to investigate the correlation of these metrics with dosimetric parameters. It could also have potential applications in the investigation of the implications of uncertainty reduction. If technical advances such as dual energy computed tomography (DECT), proton CT and improved image guidance reduce range and setup uncertainty, the relative impact of IOV on overall uncertainty becomes larger. The SA framework could complement studies such as [15–17], who have investigated the impact of range and setup margin reduction. By also including IOV into the analysis, questions such as how far the overall uncertainty can be reduced by reduction of setup and range uncertainty before IOV becomes the limiting factor could be comprehensively investigated in future studies. Similarly, the following question could be assessed: Although not explicitly accounted for in the PTV concept, it can be assumed that IOV is compensated by the margins (or, in an analogous manner in the case of robust optimization, the plan robustness settings) to a certain extent. The SA framework could help to investigate whether a CTV-to-PTV margin reduction (or reduction of plan robustness settings) justified by reduced range and setup uncertainties would lead to an unexpected increase in uncertainty of CTV coverage caused by IOV.

In conclusion, a previously presented framework for variance-based sensitivity analysis has been extended to include IOV. The approach is feasible and enables the evaluation of the combined impact of setup and range uncertainty and IOV. In a first analysis of ten patients, IOV had a relevant impact on the CTV  $D_{95\%}$  for two of these patients. This suggests that IOV could have a deteriorating effect on CTV coverage in some cases.

#### Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: CB holds research grants from Elekta, Brainlab, Viewray and C-Rad without any relation to the research described here. The remaining authors declare that they have no competing interest.

#### Acknowledgements

This project was supported by the DFG grant KA 4346/1-1 and the DFG Cluster of Excellence Munich Center for Advanced Photonics

(MAP).

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.phro.2021.11.005>.

#### References

- [1] Unkelbach J, Alber M, Bangert M, Bokrantz R, Chan TCY, Deasy JO, et al. Robust radiotherapy planning. *Phys Med Biol* 2018;63(22):22TR02. <https://doi.org/10.1088/1361-6560/aae659>.
- [2] Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. *Radiother Oncol* 2016;121(2):169–79. <https://doi.org/10.1016/j.radonc.2016.09.009>.
- [3] Lobefalo F, Bignardi M, Reggiori G, Tozzi A, Tomatis S, Alongi F, et al. Dosimetric impact of inter-observer variability for 3D conformal radiotherapy and volumetric modulated arc therapy: The rectal tumor target definition case. *Radiat Oncol* 2013; 8(1):176. <https://doi.org/10.1186/1748-717X-8-176>.
- [4] Hellebust TP, Tanderup K, Lervåg C, Fidarova E, Berger D, Malinen E, et al. Dosimetric impact of interobserver variability in MRI-based delineation for cervical cancer brachytherapy. *Radiother Oncol* 2013;107(1):13–9. <https://doi.org/10.1016/j.radonc.2012.12.017>.
- [5] Eminowicz G, Rompokos V, Stacey C, McCormack M. The dosimetric impact of target volume delineation variation for cervical cancer radiotherapy. *Radiother Oncol* 2016;120(3):493–9. <https://doi.org/10.1016/j.radonc.2016.04.028>.
- [6] Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput Phys Commun* 2010;181(2):259–70. <https://doi.org/10.1016/j.cpc.2009.09.018>.
- [7] Kamp F, Brüningk S, Cabal G, Mairani A, Parodi K, Wilkens JJ. Variance-based sensitivity analysis of biological uncertainties in carbon ion therapy. *Phys Med* 2014;30(5):583–7. <https://doi.org/10.1016/j.ejmp.2014.04.008>.
- [8] Kamp F, Wilkens JJ. Application of variance-based uncertainty and sensitivity analysis to biological modeling in carbon ion treatment plans. *Med Phys* 2019;46(2):437–47. <https://doi.org/10.1002/mp.13306>.
- [9] Zvereva A, Kamp F, Schlattl H, Zankl M, Parodi K. Impact of interpatient variability on organ dose estimates according to MIRD schema: Uncertainty and variance-based sensitivity analysis. *Med Phys* 2018;45(7):3391–403. <https://doi.org/10.1002/mp.12984>.
- [10] Hofmaier J, Dedes G, Carlson DJ, Parodi K, Belka C, Kamp F. Variance-based sensitivity analysis for uncertainties in proton therapy: A framework to assess the effect of simultaneous uncertainties in range, positioning and RBE model predictions on RBE-weighted dose distributions. *Med Phys* 2021;48(2):805–18. <https://doi.org/10.1002/mp.14596>.
- [11] Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* 2004;23(7):903–21. <https://doi.org/10.1109/TMI.2004.828354>.
- [12] Deasy JO, Blanco AI, Clark VH. CERR: A computational environment for radiotherapy research. *Med Phys* 2003;30(5):979–85. <https://doi.org/10.1118/1.1568978>.
- [13] Combs SE, Baumert BG, Bendszus M, Bozzao A, Brada M, Fariselli L, et al. ESTRO ACROP guideline for target volume delineation of skull base tumors. *Radiother Oncol* 2021;156:80–94. <https://doi.org/10.1016/j.radonc.2020.11.014>.
- [14] Kristensen I, Nilsson K, Agrup M, Belfrage K, Embring A, Haugen H, et al. A dose based approach for evaluation of inter-observer variations in target delineation. *Tech Innov Patient Support Radiat Oncol* 2017;3-4:41–7. <https://doi.org/10.1016/j.tipsro.2017.10.002>.
- [15] 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(1):56–62. <https://doi.org/10.1016/j.radonc.2016.04.038>.
- [16] Tattenberg S, Madden TM, Gorissen BL, Bortfeld T, Parodi K, Verburg J. Proton range uncertainty reduction benefits for skull base tumors in terms of normal tissue complication probability (NTCP) and healthy tissue doses. *Med Phys* 2021;48(9): 5356–66. <https://doi.org/10.1002/mp.15097>.
- [17] Wagenaar D, Kierkels RGJ, van der Schaaf A, Meijers A, Scandurra D, Sijtsma NM, et al. Head and neck IMPT probabilistic dose accumulation: Feasibility of a 2 mm setup uncertainty setting. *Radiother Oncol* 2021;154:45–52. <https://doi.org/10.1016/j.radonc.2020.09.001>.

# Application of variance-based uncertainty and sensitivity analysis to biological modeling in carbon ion treatment plans

Florian Kamp<sup>a)</sup>

Department of Radiation Oncology, Technical University of Munich, Klinikum rechts der Isar, Ismaninger Str. 22, 81675 München, Germany

Physik-Department, Technical University of Munich, James-Frank-Str. 1, 85748 Garching, Germany

Department of Radiation Oncology, University Hospital, LMU Munich, Marchioninstr. 15, 81377 München, Germany

Jan J. Wilkens

Department of Radiation Oncology, Technical University of Munich, Klinikum rechts der Isar, Ismaninger Str. 22, 81675 München, Germany

Physik-Department, Technical University of Munich, James-Frank-Str. 1, 85748 Garching, Germany

(Received 7 August 2018; revised 14 October 2018; accepted for publication 9 November 2018; published 18 December 2018)

**Purpose:** In ion beam therapy, biological models to estimate the relative biological effectiveness (RBE) and subsequently the RBE-weighted dose (RWD) are needed in treatment planning and plan evaluation. The required biological parameters as well as their dependency on ion species and ion energy can typically not be determined directly in experiments for *in vivo* situations. For that reason they are often derived from *in vitro* data and biological modeling and subject to large uncertainties. We present a model-independent Monte Carlo (variance-) based uncertainty and Sensitivity Analysis (SA) approach to quantify the impact of different input uncertainties on a simulated carbon ion treatment plan.

**Method:** The influences of different input uncertainties are examined by variance-based SA methods. In this Monte Carlo approach, a function is evaluated  $10^3$ – $10^5$  times. For each of those runs, all inputs are changed simultaneously, using random numbers according to their associated uncertainties. Variance-based statistic formalisms then rank the input parameter/uncertainty pairs according to their impact on the result of the function. The method of SA includes an uncertainty analysis and was applied to a two-field spot scanning carbon ion treatment plan for two commonly used biological models and two representative tissue parameter sets.

**Results:** Based on an exemplary patient case, the application of variance-based SA for biological measures, relevant in (carbon) ion therapy, is demonstrated. A voxel-wise calculation for  $2.9 \cdot 10^5$  voxels takes  $\sim 6$  h. A structure-based SA, which adds an uncertainty band to a RWD-volume histogram (RW-DVH) and shows how to decrease the uncertainty in the most effective way, can be calculated in 0.1–1.5 h (depending on the size of the structure). The uncertainties in RBE, RWD or RW-DVH are broken down to the impact of different uncertainties in the (biological) model input. Biological uncertainties have a higher impact on the resulting RBE and RWD than uncertainties in the physical dose. Excluding the physical dose from the SA only slightly decreased the overall uncertainty, emphasizing the necessity to include biological uncertainties into treatment plan evaluation.

**Conclusion:** Variance-based SA is a powerful tool to evaluate the impact of uncertainties in (carbon) ion therapy. The number of input parameters that can be examined at once is only limited by computation time. A Monte Carlo-derived, comprehensive uncertainty quantification and a corresponding sensitivity analysis are implemented and provide new information for treatment plan evaluation. A possible future application is a SA-based biologically robust treatment plan optimization using the additional uncertainty information as presented here. © 2018 American Association of Physicists in Medicine [https://doi.org/10.1002/mp.13306]

Key words: carbon ion radiotherapy, relative biological effectiveness (RBE), uncertainty and sensitivity analysis, uncertainty propagation

## 1. INTRODUCTION

In ion beam therapy, biological models to estimate the relative biological effectiveness (RBE) are frequently used in treatment planning and plan evaluation. In the context of the linear-quadratic (LQ) model,<sup>1</sup> the RBE depends on biological parameters ( $\alpha_p$  and  $\beta_p$ ) for particles/ions as well as for the

reference x-ray radiation ( $\alpha_x$  and  $\beta_x$ ) and the dose per fraction ( $d$ ). The needed biological parameters as well as their dependency on ion species and ion energy typically cannot be determined directly in experiments for *in vivo* situations. They are often derived from *in vitro* data in combination with biological modeling and are subject to large uncertainties. For a comprehensive treatment plan evaluation and

optimization, it is therefore necessary to estimate the resulting uncertainties in e.g., RBE or RBE-weighted dose (RWD) caused by the uncertainties of the relevant input parameters. In general RBE and hence also uncertainties in RBE predictions are more important in carbon ion therapy than in proton therapy. Typical RBE values inside the planning target volume (PTV) range between 3 and 5 for carbon ion therapy and 1.1 for proton therapy.

Currently, a lot of effort is invested into the quantification and subsequent minimization of dose-related uncertainties for proton and ion therapy.<sup>2–4</sup> This includes both the quantification of these physical uncertainties<sup>5–9</sup> and how to account for them in the optimization process.<sup>10–16</sup> Little is known about the actual size and influence of uncertainties in RBE predictions. The problem is that the data for estimating uncertainties of RWD in the clinics is scarce, ambiguous, and contains biases (e.g., *in vitro* vs *in vivo* differences). Therefore, clinical decisions are commonly only based on an opinion of an expert (and are usually not systematic and quantitative).

Computationally simple sensitivity analysis approaches have already been used in ion beam therapy, especially to address the biological effects of carbon ion beams (e.g., Refs. [17–19]). The authors reported significant variations due to systematic changes in individual parameters of the Local Effect Model as well as  $\alpha_x$  and  $\beta_x$ . These earlier sensitivity analysis approaches used for biological model predictions in (carbon) ion therapy were carried out thoroughly by changing one or two input parameters at a time by up to  $\pm 50\%$ . The authors reported proportionally lower changes in the RBE compared to the changed input and that uncertainties in general decrease in Spread-out Bragg Peaks (SOBP) compared to monoenergetic beams. Changing one factor at a time is a commonly used, computational inexpensive method to quantify the effect of uncertain inputs to models. The method itself is limited as it comes without the possibility to quantify an overall uncertainty considering all input parameters at once and is hence potentially ignoring interplay effects between different inputs. In order to overcome these limitations, Kamp et al.<sup>20</sup> showed a first application of variance-based SA for carbon ion therapy. This proof-of-concept study was limited to a simplified one-dimensional (1D) SOBP setting without a repeated full execution of the biological modeling process. In this study, variations in RBE and RWD were computed assigning uncertainties to three parameters:  $\alpha_p$ ,  $\beta_p$  and  $d$  (whereas  $\alpha_x$  and  $\beta_x$  had to be kept constant).

It becomes obvious that for both, comprehensive treatment plan evaluation and their biologically robust optimization, two main challenges need to be addressed. First, a comprehensive uncertainty quantification is needed, considering as many input uncertainties as feasible at once as well as their potential interplay. Second, a sensitivity measure is needed to provide additional information on the relative impact of different input uncertainties on the resulting comprehensive uncertainty.

We employed a statistical method called variance-based SA in order to quantify uncertainties of RBE and RWD distributions. The employed variance-based approach randomly

varies all input parameters at once. In this Monte Carlo approach, a comprehensive uncertainty in RBE or RWD is modeled which includes all specified input uncertainties as well as their interplay effects. With a statistical ranking, the so-called SA, the simulated variance in the result can be broken down into the impact of the different input uncertainties.

This means that the work of Kamp et al.<sup>20</sup> was extended from a 1D SOBP in water to a 3D carbon ion treatment plan of a real patient case. In addition, a very fast biological model was implemented. This facilitates the extension of the analyses to changes in all input parameters of RBE predictions. The step from 1D to 3D is accompanied by the possibility to evaluate structure-based measures as, e.g., dose–volume histograms and RWD-quantiles. To cover a broader range of possible  $\alpha_x/\beta_x$  ratios, two representative values ( $\alpha_x/\beta_x = 2$  and 9.2 Gy) are chosen in this manuscript (in Kamp et al.<sup>20</sup> only one  $\alpha_x/\beta_x = 6.4$  Gy was used).

## 2. MATERIALS AND METHODS

### 2.A. Research treatment planning system

We added a carbon ion treatment planning extension to the Computational Environment for Radiation Research platform (CERR).<sup>21–24</sup> This extension includes 3D, voxel-wise dose calculation, treatment plan optimization based on the biological effect and the calculation of RBE, RWD, dose-weighted LET, as well as dose-weighted  $\alpha_p$  and  $\beta_p$ . The reference radiosensitivity parameters  $\alpha_x$  and  $\beta_x$  can be specified for every voxel. The implemented multi-field optimization uses either squared differences in the biological effect<sup>25</sup> or the concept of equivalent uniform effect as reported by Brüningk et al.<sup>23</sup>

The calculations are based on a pencil beam, spot scanning algorithm. The necessary fragmentation spectra were simulated according to Parodi et al.<sup>26</sup> for a clinically representative beam line and tabulated. This includes spectra of the fully ionized elements with atomic numbers  $Z_{ion} = 1$  to 6.

### 2.B. Prediction of RBE

In the framework of the LQ model, isoeffective RBE can be calculated as a function of five input parameters.

$$\begin{aligned} RBE(\alpha_p, \beta_p, \alpha_x, \beta_x, d) \\ = \frac{1}{2\beta_x d} \left( -\alpha_x + \sqrt{\alpha_x^2 + 4\beta_x(\alpha_p d + \beta_p d^2)} \right) \end{aligned} \quad (1)$$

In order to predict  $\alpha_p$  and  $\beta_p$  for a given set of x-ray reference parameters ( $\alpha_x$  and  $\beta_x$ ) biological model predictions were implemented into CERR. In this work, we used the repair-misrepair-fixation (RMF) model<sup>24</sup> and the Local Effect Model 1 (LEM1).<sup>27</sup>

#### 2.B.1. Repair-misrepair-fixation (RMF) model

The mechanistic RMF model was developed by Carlson et al.<sup>28</sup> here it is implemented based on double-strand break



(DSB) inductions from the Monte Carlo damage simulation (MCDS).<sup>29–31</sup> The RMF model was implemented based on MCDS version 3.10 and a default cell nuclear diameter of 5  $\mu\text{m}$ .

The RMF model predictions for ion therapy have been shown and discussed in several publications.<sup>24,28,32–34</sup> We focus here on the properties of the RMF model that are favorable for the presented variance-based SA. Kamp et al.<sup>33</sup> introduced a decoupling approach for the RMF model. The dose-weighted  $\alpha_p$  and  $\beta_p$  in every voxel  $i$  can be expressed as

$$\alpha_{p,i} = \alpha_{x,i} \cdot c_{1,i} + \beta_{x,i} \cdot c_{2,i} \quad (2)$$

$$\beta_{p,i} = \beta_{x,i} \cdot (c_{1,i})^2 \quad (3)$$

with  $c_{1,i}$  and  $c_{2,i}$  as introduced by Kamp et al.<sup>33</sup>  $c_{1,i}$  and  $c_{2,i}$  combine physical and biological beam properties, such as particle type and spectra, stopping power, frequency-mean specific energy, and the simulated DSB induction to two values for every voxel.  $c_{1,i}$  and  $c_{2,i}$  are explicitly decoupled from the reference radiosensitivity parameters  $\alpha_{x,i}$  and  $\beta_{x,i}$ .

Practically this means that within the framework of the RMF model, predictions can be adapted very fast for changes in  $\alpha_x$  and  $\beta_x$ . The influence of a random change of  $\alpha_{x,i}$  and  $\beta_{x,i}$  can be calculated extremely fast for all voxels as it is a simple vector multiplication and summation. In the framework of the RMF model, RBE reduces to a function of the frequently used  $\alpha_x/\beta_x$  ratio and  $c_1$ ,  $c_2$ , and  $d$ .

$$\begin{aligned} \text{RBE}_{\text{RMF}} \left( \frac{\alpha_x}{\beta_x}, c_1, c_2, d \right) \\ = \frac{1}{2d} \left( -\frac{\alpha_x}{\beta_x} \sqrt{\left( \frac{\alpha_x}{\beta_x} \right)^2 + 4d \left( c_1 \frac{\alpha_x}{\beta_x} + c_2 + c_1^2 d \right)} \right) \end{aligned} \quad (4)$$

This equation is derived by combining Eq. (1) with Eqs. (2) and (3) as shown by Kamp et al.<sup>33</sup>

### 2.B.2. Local effect model 1 (LEM1)

For comparison, we implemented the LEM1<sup>27</sup> predictions to the prescribed extension of CERR. LEM1 is currently used for the clinical treatment planning of patients.<sup>35</sup> Tabulated dependencies of  $\alpha_p$  and  $\beta_p$  on particle type and energy are taken from the LEM1 implementations on the INFN/I-SEE homepage (<http://totlxl.to.infn.it/lem/>, accessed 2015, currently not available). Note that a fast recalculation, comparable to the  $c_1$ – $c_2$  formalism above, is not possible using LEM. Every time  $\alpha_x$  or  $\beta_x$  are changed, LEM1 needs to be rerun to simulate a new set of corresponding  $\alpha_p$  and  $\beta_p$ .

### 2.C. Sensitivity analysis

We adopted the Factor Prioritization approach as described by Saltelli et al.<sup>36</sup> for carbon ion treatment plan evaluation for 3D patient cases. This SA approach ranks uncertainties of different inputs according to their influence on an uncertainty in the result. The Factor Prioritization is

achieved by a Monte Carlo approach, followed by a variance-based ranking of the influence of different input.

#### 2.C.1. Variance-based sensitivity analysis

The Factor Prioritization approach by Saltelli et al.,<sup>36</sup> combines a Monte Carlo sampling for the uncertainties of the inputs with a variance-based evaluation. This means that a function is evaluated  $n_{\text{run}}$  times, each time with a different set of randomly changed input parameters. The input is changed randomly according to the individually assigned random number distributions, accounting for the associated uncertainties. A variance-based statistic ranks the uncertainty in the different inputs according to their impact on the result. The sensitivity  $S_k$  on the  $k$ -th input parameter  $X_k$  is calculated as

$$S_k = \frac{\text{var}(\text{mean}(Y|X_k \approx \text{const}))}{\text{var}(Y)}, \quad (5)$$

with  $Y$  being the result vector of  $n_{\text{run}}$  function evaluations. Following the description in Kamp et al.,<sup>20</sup> the numerator is calculated by first sorting the  $(X_k, Y)$  pairs by increasing  $X_k$ . The sorted pairs are then divided into  $n_{\text{par}}$  partitions, each containing  $n_{\text{run}}/n_{\text{par}}$  entries with increasing  $X_k$ . The mean values are taken over these partitions, the variance then from the  $n_{\text{par}}$  mean values. Extensive testing showed that for the here discussed study the combination of  $n_{\text{run}} = 131\,072 = 2^{17}$  and  $n_{\text{par}} = 256 = 2^8$  was numerically stable. The implemented approach provides not only an uncertainty of the result (standard deviation  $\sigma_Y = \sqrt{\text{var}(Y)}$ ) but also the corresponding sensitivity information. These first-order sensitivity values range from  $S = 1$  (only influential input) to  $S = 0$  (no influence). If  $\sum_k S_k = 1$ , it can be inferred that no higher order sensitivity values (cross terms of two or more inputs) are present.<sup>36</sup>

#### 2.C.2. Adaption to carbon ion therapy treatment planning

Computation time and memory are the most crucial factors for the implementation of the SA. For instance, the RBE and RWD of the presented patient case with  $2.9 \cdot 10^5$  relevant voxels need to be evaluated  $n_{\text{run}} = 131\,072$  times in every voxel, then sorted and inserted into Eq. (5). In order to achieve reasonable computation times with optimal memory usage, several parallelization steps are implemented.

The first step is to use the same relative random numbers for all voxels. This means, the relative changes are randomly sampled for every input but the same in all voxels. For example, can  $n_{\text{run}}$  changes of  $\alpha_{x,i}$  be derived by a simple multiplication  $\alpha_{x,i} \cdot \Delta_{\alpha_x}$  for all voxels  $i$ . Here  $\Delta_k$  are vectors of length  $n_{\text{run}}$  with relative changes ( $\Delta_k = 1$  is equal to the unchanged value). Note that all  $\Delta_k$  are explicitly independent of the voxel  $i$ . This not only ensures spatially continuous values in every run but also reduces the amount of sorting needed for the SA to the absolute minimum. In this way, the random numbers

need to be sorted only once per input and the sorting indices can be applied to all voxels.

For the second parallelization step, two different approaches need to be considered. The first can be applied to voxel-based measures as for instance spatial distribution of  $RWD_i$ ,  $RBE_i$ ,  $\alpha_{p,i}$ , or  $\beta_{p,i}$ . The second is applied for structure-based measures as, e.g.,  $RWD_{50\%}$ ,  $RWD_{2\%}$ , or  $RWD_{98\%}$  where  $RWD_{X\%}$  denotes the minimal RBE-weighted dose received by  $X\%$  of the structure's volume.

Voxel-based measures are calculated for  $n_{vox}$  in parallel to optimize memory usage and to speed up the calculation. Typical values for  $n_{vox}$  lie between  $10^2$  and  $10^4$  and scale with  $n_{run}^{-1}$ . The resulting uncertainty ( $\sigma_{Y,i}$ ) and sensitivity ( $S_{k,i}$ ) maps can be displayed similar to dose maps, as colorwash overlay on the CT image of the patient.

Structure-based measures need the information in all voxels of the structure at once. They can be calculated in parallel for a subgroup of runs  $n_{sub}$ . Typical values for the size of these subgroups lie between  $10^2$  and  $10^3$  and are inverse proportional to the number of voxels in a structure. The result is one sensitivity value per structure and input parameter combination.

### 2.C.3. Used SA settings and assigned random number distributions

The presented SAs were done based on  $n_{run} = 131\,072$  subdivided into  $n_{par} = 256$  partitions. Table I summarizes the evaluated measures, their dependencies, and the relative uncertainty assigned to the different inputs. The table shows relative standard deviations  $\sigma_k$  which are used to generate normally distributed random numbers with an expectation value of 0. The relative changes  $\Delta_k$  are obtained by adding +1 to the resulting random numbers. We chose normally distributed random numbers, as this distribution is commonly used to report uncertainties. The relative standard deviations were chosen following the values reported by Weyrather et al.<sup>35</sup> and Friedrich et al.<sup>37</sup> for the biological parameters. The uncertainty in the dose is a rough estimation of the achieved relative uncertainty considering clinical dose delivery. The SA itself can in general be executed with any distribution, whichever describes the uncertainty in the input best. Note that the comprehensive SA of columns 1–3 in Table I can only be done for RMF model predictions at the moment. The reduced SA, mentioned in the fourth column, can be done for LEM1 as well. This SA is reduced, as  $\alpha_p(\alpha_x, \beta_x)$  and  $\beta_p(\beta_x, c_1)$  are not changed accordingly. The reduced SA needed to be introduced as the current database does not support arbitrary changes in the LEM1 input ( $\alpha_x$  and  $\beta_x$ ). Any change in  $\alpha_x$  and  $\beta_x$  in the LEM1 would require a new model execution to simulate  $\alpha_p$  and  $\beta_p$ . For the implemented SA, this would mean  $n_{run} = 131\,072$  LEM1 executions which are computationally not feasible.

Note that instead of  $RBE(\alpha_p, \beta_p, d)$  as reported in Kamp et al.<sup>20</sup>, we simulate  $RBE(\alpha_x, \beta_x, d)$  in the reduced SA. Due to the involved sorting in the calculation of  $S_k$ , the SA cannot be executed for correlated or dependent input. As  $\alpha_p$  and  $\beta_p$

TABLE I. Evaluated functions and the assigned uncertainties in their inputs. The uncertainties are modeled with normal distributions.  $^*\sigma_{\alpha_x/\beta_x} = 15\%$  is an approximated uncertainty in  $\alpha_x/\beta_x$ , assuming  $\sigma_{\alpha_x} = 10\%$  and  $\sigma_{\beta_x} = 10\%$ .  $^{\#}0.001\%$  represents a very small uncertainty in the dose.

		$RBE_{RMF}$ ( $\alpha_x/\beta_x, c_1, c_2, d$ )	$RBE_{RMF/LEM}$ ( $\alpha_x, \beta_x, d$ )
		$RWD_{RMF}$ ( $\alpha_x/\beta_x, c_1, c_2, d$ )	$RWD_{RMF/LEM}$ ( $\alpha_x, \beta_x, d$ )
$\alpha_p(\alpha_x, \beta_x, c_1, c_2)$	$\beta_p(\beta_x, c_1)$		
$\sigma_{\alpha_x} = 10\%$	$\sigma_{\beta_x} = 10\%$	$\sigma_{\alpha_x/\beta_x} = 15\%^*$	$\sigma_{\alpha_x} = 10\%$
$\sigma_{\beta_x} = 10\%$	$\sigma_{c_1} = 10\%$	$\sigma_{c_1} = 10\%$	$\sigma_{\beta_x} = 10\%$
$\sigma_{c_1} = 10\%$		$\sigma_{c_2} = 10\%$	$\sigma_d = 5\%$
$\sigma_{c_2} = 10\%$		$\sigma_d = 5\%$ and $0.001\%^{\#}$	

are dependent on the same variables, a reduced SA modeling  $RBE(\alpha_x, \beta_x, d)$  is better suited for comparisons with the comprehensive SA.

## 3. RESULTS

We demonstrate and discuss the presented SA on the basis of an astrocytoma patient previously treated with photons. The nominal treatment planning result is obtained by a simultaneous multi-field optimization. Spatially constant  $\alpha_x$  and  $\beta_x$  were assigned throughout the patient. Two representative tissue parameter sets were chosen. One with a small  $\alpha_x/\beta_x = 2$  Gy ( $\alpha_x = 0.1$  Gy<sup>-1</sup> and  $\beta_x = 0.05$  Gy<sup>-2</sup>, representing chordomas of the skull base<sup>38</sup>) and one with a large  $\alpha_x/\beta_x = 9.2$  Gy ( $\alpha_x = 0.184$  Gy<sup>-1</sup> and  $\beta_x = 0.02$  Gy<sup>-2</sup>, for Chinese hamster fibroblasts V79<sup>39</sup>). The treatment plan was optimized on a RWD of 3 Gy(RBE) in the PTV. The later discussed Figs. 1 and 2 display different quantities for the same RMF model-based carbon ion treatment plan calculated with the smaller  $\alpha_x/\beta_x = 2$  Gy.

All simulations were performed in a computational environment equipped with 8 cores (2.66 GHz) and 32 GB random-access memory within the Matlab 2014a (The MathWorks, Inc., Natick, MA) software package.

In the following, the different uncertainty and SA results are reported. A common result is that, even though several inputs are changed simultaneously with  $\sigma = 5\% - 15\%$ , the relative change in the examined RBE and RWD is smaller than 10%.

### 3.A. Sensitivity maps for RMF model predictions

Figure 1 shows the resulting sensitivity maps for  $\alpha_p(\alpha_x, \beta_x, c_1, c_2)$  and  $\beta_p(\beta_x, c_1)$  modeled for a spatially constant  $\alpha_x/\beta_x = 2$  Gy ( $\alpha_x = 0.1$  Gy<sup>-1</sup> and  $\beta_x = 0.05$  Gy<sup>-2</sup>). The modeled, unchanged distributions of  $\alpha_p$  and  $\beta_p$  are shown in panels a and b. Both increase with increasing LET. The spatial distribution of LET is shown in panel c. The resulting standard deviations of  $\alpha_p$  and  $\beta_p$  (panels d and e, respectively) show the same behavior as  $\alpha_p$  (panel a) and  $\beta_p$  (panel b), respectively. Their relative standard deviation (derived by “dividing panel d by panel a” or “panel e by panel b”) ranges from 10.1% to 12.4% for  $\alpha_p$  and is spatially constant at 22.6%



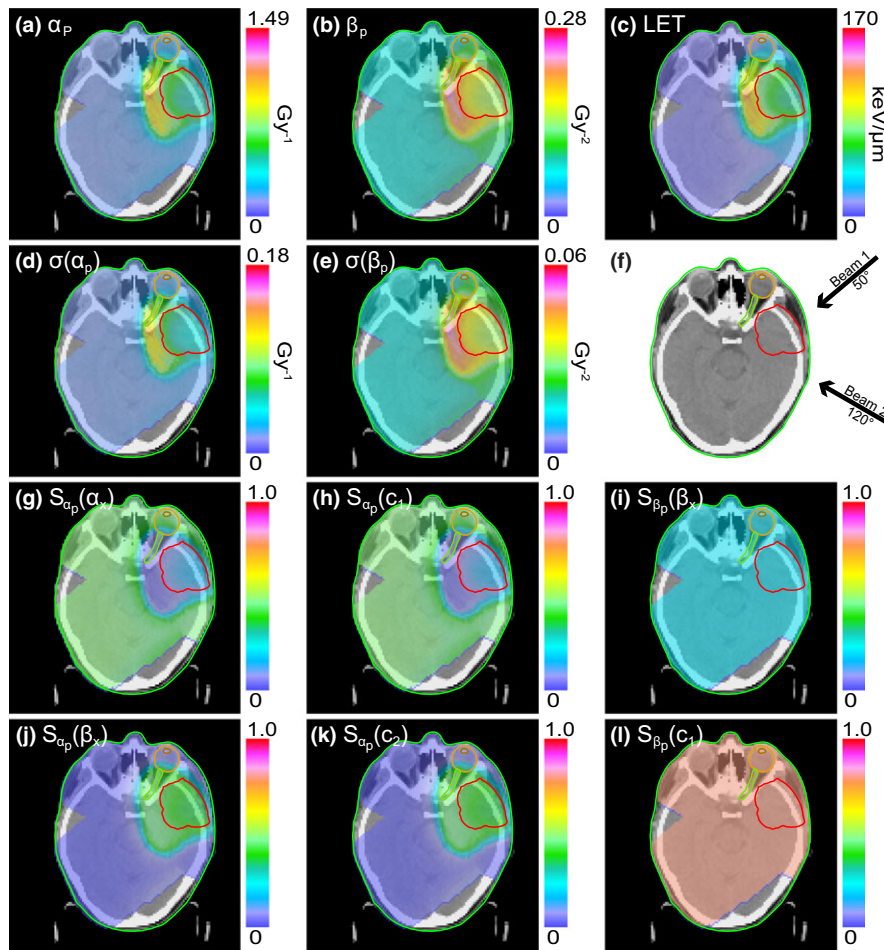


FIG. 1. Sensitivity maps for  $\alpha_p(\alpha_x, \beta_x, c_1, c_2)$  and  $\beta_p(\beta_x, c_1)$ . The panels show the modeled  $\alpha_p$  and  $\beta_p$  together with their resulting standard deviation panels (a), (b), (d) and (e). Panels (c) and (f) show a LET map and the simulated beam geometry. The six sensitivity maps, four for  $\alpha_p$  and two for  $\beta_p$ , are shown in panels (g–l). The underlying treatment plan, including a spatial distribution of  $c_1$  and  $c_2$ , can be seen in Fig. 2. The planning target volume (PTV) is marked in red, the left optical nerve in green together with the left eyeball (orange) and the left lens (brown). [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

for  $\beta_p$ . Panels g, h, j, and k show the corresponding sensitivity maps for  $\alpha_p$ .  $S_{\alpha_p}(\alpha_x) = S_{\alpha_p}(c_1)$  and  $S_{\alpha_p}(\beta_x) = S_{\alpha_p}(c_2)$  while the spatial distribution of  $S_{\alpha_p}(\alpha_x)$  and  $S_{\alpha_p}(c_1)$  are different to  $S_{\alpha_p}(\beta_x)$  and  $S_{\alpha_p}(c_2)$ .  $\alpha_p$  is more dependent on changes in  $\alpha_x$  and  $c_1$  for voxels with lower LET, while uncertainties in  $\beta_x$  and  $c_2$  have a higher impact for high LET regions. The sensitivity maps for  $\beta_p$  are shown in panels i and l.  $S_{\beta_p}(\beta_x) \approx 0.20$  and  $S_{\beta_p}(c_1) \approx 0.80$  are spatially constant. The uncertainty in  $c_1$  has a higher impact on  $\beta_p$  than the uncertainty in  $\beta_x$ .

Figure 2 displays the resulting sensitivity maps for  $RWD_{RMF}(\alpha_x/\beta_x, c_1, c_2, d)$ . The plan was optimized on 3 Gy (RBE) in the PTV using a spatially constant  $\alpha_x/\beta_x = 2$  Gy. The optimized RWD (panels a) is displayed together with its uncertainty (panel d). The LET (panel b) and the beam geometry (panel e) are shown together with the four unchanged input parameter maps (panel c, g–i). The four corresponding sensitivity maps are shown in panel f, j–l. The shown voxel-wise SA for  $2.9 \cdot 10^5$  voxels (Figs. 1 and 2 and Fig. S1) took 6 h in total.

Several relations can be observed. The impact of changing  $d$  is almost constant spatially  $S_{RWD}(d) \approx 0.3$ . The  $S_{RWD}(\alpha_x/\beta_x)$  is small throughout the patient except in the

region 1 to 2.5 cm distal to the PTV. In this region,  $S_{RWD}(\alpha_x/\beta_x)$  reaches values up to 0.5. This means that the uncertainty in  $(\alpha_x/\beta_x)$  has the greatest impact on the uncertainty in RWD in this region. The sensitivity maps of the two biological modeling parameters  $c_1$  and  $c_2$  show opposite behavior:  $S_{RWD}(c_1)$  is up to 0.7 in the PTV and in the fragmentation tail far from the PTV. In the region 1 to 2.5 cm distal to the PTV it drops to very small values  $S_{RWD}(c_1) \leq 0.05$ , whereas  $S_{RWD}(c_2)$  is small everywhere except in the PTV  $S_{RWD}(c_2) \approx 0.15$  and the mentioned region from 0 to 2.5 cm distal to the PTV ( $S_{RWD}(c_2) \approx 0.4$ ). The high values in  $S_{RWD}(\alpha_x/\beta_x)$  and  $S_{RWD}(c_2)$  and the corresponding small values of  $S_{RWD}(c_1)$  are located in voxels with high LET.

For the sake of completeness the sensitivity maps for  $RBE_{RMF}(\alpha_x/\beta_x, c_1, c_2, d)$  are shown in the supplementary material, available online (Fig. S1).

### 3.B. Structure-based sensitivity analysis for RMF model predictions

Figure 3 displays RW-DVHs together with the uncertainty information for the two representative  $\alpha_x/\beta_x$ , modeled with

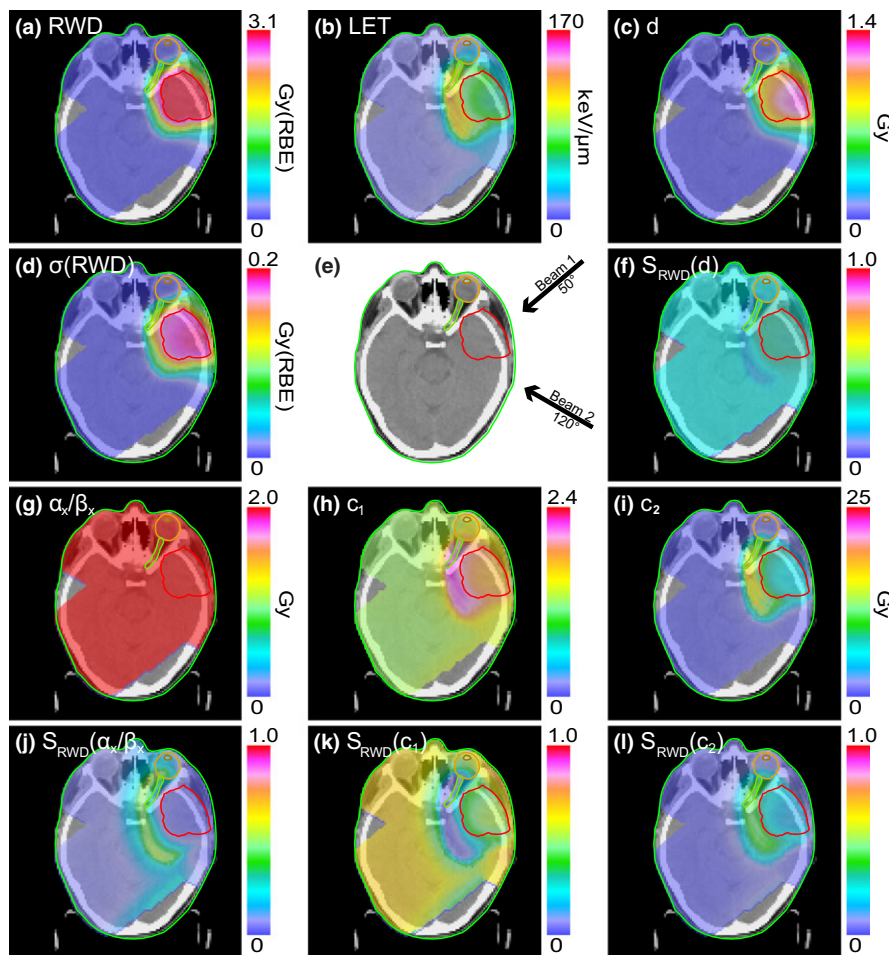


FIG. 2. Sensitivity maps for  $RWD_{RMF}(\alpha_x/\beta_x, c_1, c_2, d)$ . The optimized RWD (a) is displayed together with its uncertainty (panel d). The LET (panel b) and the beam geometry (e) are shown together with the four unchanged input parameter maps (panel c, g–i). The four sensitivity maps are shown in panels (f, j–l). The planning target volume (PTV) is marked in red, the left optical nerve in green together with the left eyeball (orange), and the left lens (brown). [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

and without an uncertainty in  $d$ . The uncertainty is added to the plots, using the variation in the  $n_{run}$  RW-DVHs from the SA. Table II summarizes resulting  $RWD_{98\%}$ ,  $RWD_{50\%}$ , and  $RWD_{2\%}$  representing the low, median, and high RWD, respectively. Together with their uncertainty ( $\sigma_{RWD_{X\%}}$ ), the results of the SA are reported. The sensitivity values break down the uncertainty for a relative volume in the RW-DVH into the different input uncertainties. The RW-DVHs of the treatment plan simulated with  $\alpha_x/\beta_x = 2$  Gy and  $\alpha_x/\beta_x = 9.2$  Gy are very similar. The resulting uncertainty is larger for  $\alpha_x/\beta_x = 9.2$  Gy. Assigning a very small uncertainty to  $d$ , decreases the uncertainty by 0.03–0.04 Gy for the PTV and by 0.00–0.03 Gy for the left optical nerve.

Comparing the sensitivity values for the PTV in Table II, it can be observed that  $c_1$  becomes more important for  $\alpha_x/\beta_x = 9.2$  Gy, whereas the uncertainty of  $c_2$  is less important. Comparison of the  $S_{X\%}$  values for the PTV and the left optical nerve demonstrates that different structures can have very different SA results. Depending on the location of the structure, different input parameter uncertainties have a different impact.  $\alpha_x/\beta_x$  and  $c_2$  are more important for the left

optic nerve than for the PTV. For  $c_1$  and  $d$ , it is the other way. This is valid for both cell lines and was already observed in the axial slices of the 3D sensitivity maps in Fig. 2.  $S_{X\%}(\alpha_x/\beta_x) + S_{X\%}(d) + S_{X\%}(c_1) + S_{X\%}(c_2) \approx 1$  for all shown structures, cell lines, and evaluated  $RWD_{X\%}$ . The computation time was 0.1 h for the left optical nerve ( $10^2$  voxels, representing a very small structure  $<1$  cm $^3$ ) and 1.5 h for the PTV ( $3.5 \cdot 10^5$  voxels, representing a large structure  $\approx 270$  cm $^3$ ).

### 3.C. Reduced SA for different biological models

Table III summarizes the SA results for the reduced SA. Here RWD was calculated as a function of  $\alpha_x$ ,  $\beta_x$ , and  $d$ , whereas  $\alpha_p$  and  $\beta_p$  were kept constant. Independent of the chosen  $\alpha_x/\beta_x$ , the RMF model as well as the LEM1 based RWD are very similar in the PTV. The LEM1 predicts slightly increased  $RWD_{X\%}$  for the left optical nerve. The corresponding uncertainties  $\sigma_{RWD_{X\%}}$  in PTV and optical nerve are almost identical for the different biological models. For both biological models, the

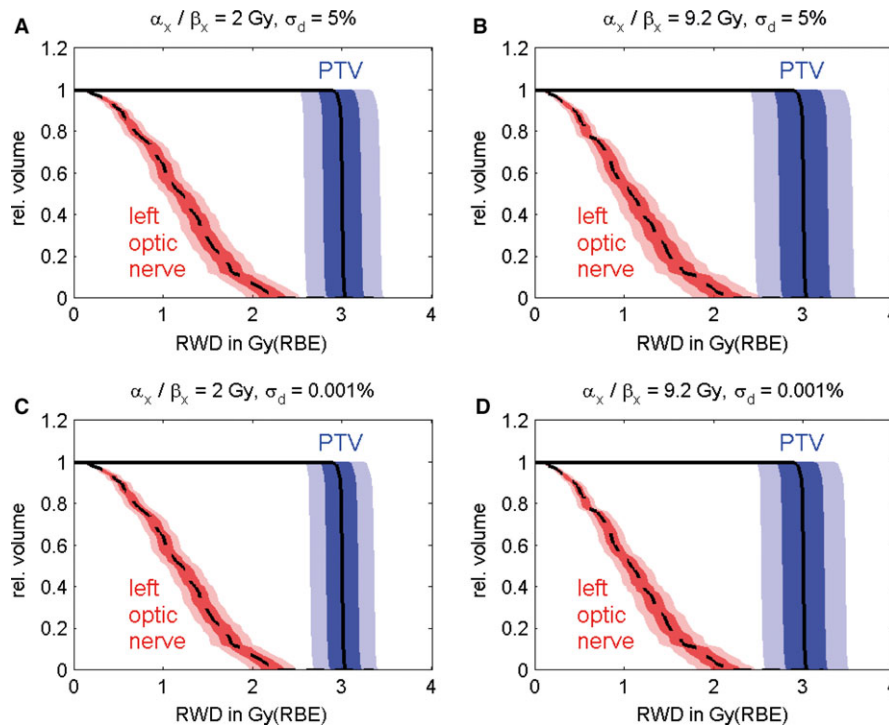


FIG. 3. RW-DVH for the PTV and the left optic nerve for two representative  $\alpha_x/\beta_x$ : 2 Gy panels (a) and (c) and 9.2 Gy panels (b) and (d). Panels (a) and (b) display the result of a SA with a modeled uncertainty in the physical dose ( $\sigma_d = 5\%$ ), for the results in panels (c) and (d) the uncertainty in the dose was neglected ( $\sigma_d = 0.001\%$ ). The black RW-DVH lines denote the result of the unchanged optimization result. The uncertainties in the RW-DVH are indicated as  $\pm \sigma$  (68.3% probability to lie inside the dark area) and  $\pm 2\sigma$  (95.4% probability to lie inside the light and dark area). [Color figure can be viewed at wileyonlinelibrary.com]

TABLE II.  $RWD_{2\%}$ ,  $RWD_{50\%}$  and  $RWD_{98\%}$ , and their uncertainties ( $\sigma_{RWD_{X\%}}$ ). The result of the sensitivity analysis is reported for the PTV and left optic nerve, with and without a simulated uncertainty in the physical dose  $d$ . Two representative  $\alpha_x/\beta_x$  values were evaluated. The corresponding RW-DVHs are displayed in Fig. 3.

$\alpha_x/\beta_x = 2$ Gy	PTV, $\sigma_d = 5\%$				$\alpha_x/\beta_x = 9.2$ Gy	PTV, $\sigma_d = 5\%$					
	$RWD_{X\%} \pm \sigma_{RWD_{X\%}}$ [Gy(RBE)]	$\frac{\alpha_x}{\beta_x}$	$d$	$c_1$		$c_2$	$RWD_{X\%} \pm \sigma_{RWD_{X\%}}$ [Gy (RBE)]	$\frac{\alpha_x}{\beta_x}$	$d$	$c_1$	$c_2$
$RWD_{2\%} = 3.04 \pm 0.21$	$S_{2\%}$	0.04	0.31	0.43	0.21	$RWD_{2\%} = 3.03 \pm 0.27$	$S_{2\%}$	0.06	0.27	0.62	0.06
$RWD_{50\%} = 2.99 \pm 0.21$	$S_{50\%}$	0.03	0.31	0.49	0.17	$RWD_{50\%} = 2.99 \pm 0.28$	$S_{50\%}$	0.04	0.26	0.66	0.04
$RWD_{98\%} = 2.92 \pm 0.20$	$S_{98\%}$	0.05	0.31	0.42	0.23	$RWD_{98\%} = 2.93 \pm 0.26$	$S_{98\%}$	0.05	0.26	0.63	0.06
	PTV, $\sigma_d = 0.001\%$					PTV, $\sigma_d = 0.001\%$					
$RWD_{2\%} = 3.04 \pm 0.17$	$S_{2\%}$	0.06	0.00	0.63	0.30	$RWD_{2\%} = 3.04 \pm 0.23$	$S_{2\%}$	0.08	0.00	0.84	0.08
$RWD_{50\%} = 2.99 \pm 0.18$	$S_{50\%}$	0.04	0.00	0.72	0.24	$RWD_{50\%} = 3.00 \pm 0.24$	$S_{50\%}$	0.05	0.00	0.90	0.06
$RWD_{98\%} = 2.92 \pm 0.17$	$S_{98\%}$	0.06	0.00	0.61	0.32	$RWD_{98\%} = 2.93 \pm 0.23$	$S_{98\%}$	0.07	0.00	0.86	0.08
$\alpha_x/\beta_x = 2$ Gy	Optic nerve, $\sigma_d = 5\%$				$\alpha_x/\beta_x = 9.2$ Gy	Optic nerve, $\sigma_d = 5\%$					
	$RWD_{X\%} \pm \sigma_{RWD_{X\%}}$ [Gy (RBE)]	$\frac{\alpha_x}{\beta_x}$	$d$	$c_1$		$c_2$	$RWD_{X\%} \pm \sigma_{RWD_{X\%}}$ [Gy (RBE)]	$\frac{\alpha_x}{\beta_x}$	$d$	$c_1$	$c_2$
$RWD_{2\%} = 2.19 \pm 0.15$	$S_{2\%}$	0.14	0.28	0.17	0.42	$RWD_{2\%} = 2.10 \pm 0.19$	$S_{2\%}$	0.15	0.27	0.47	0.12
$RWD_{50\%} = 1.20 \pm 0.10$	$S_{50\%}$	0.29	0.22	0.08	0.42	$RWD_{50\%} = 1.07 \pm 0.10$	$S_{50\%}$	0.25	0.25	0.37	0.14
$RWD_{98\%} = 0.18 \pm 0.02$	$S_{98\%}$	0.49	0.16	0.09	0.25	$RWD_{98\%} = 0.15 \pm 0.02$	$S_{98\%}$	0.19	0.23	0.47	0.08
	Optic nerve, $\sigma_d = 0.001\%$					Optic nerve, $\sigma_d = 0.001\%$					
$RWD_{2\%} = 2.19 \pm 0.13$	$S_{2\%}$	0.19	0.00	0.23	0.58	$RWD_{2\%} = 2.10 \pm 0.16$	$S_{2\%}$	0.20	0.00	0.64	0.16
$RWD_{50\%} = 1.20 \pm 0.09$	$S_{50\%}$	0.36	0.00	0.11	0.53	$RWD_{50\%} = 1.07 \pm 0.09$	$S_{50\%}$	0.32	0.00	0.49	0.19
$RWD_{98\%} = 0.18 \pm 0.02$	$S_{98\%}$	0.58	0.00	0.10	0.30	$RWD_{98\%} = 0.15 \pm 0.01$	$S_{98\%}$	0.24	0.00	0.62	0.10

uncertainty  $\sigma_{RWD_{X\%}}$  is larger for the larger  $\alpha_x/\beta_x$ . Figure S2 shows the corresponding RW-DVHs plotted analog to Fig. 3.

For both evaluated  $\alpha_x/\beta_x$  values and for both structures,  $S_{X\%}^{RMF}(d)$  was greater than  $S_{X\%}^{LEM1}(d)$ , indicating that an uncertainty in the dose has a higher impact on the resulting

variance in the RWD predictions. For both biological models and both structures, it can be found that  $S_{X\%}^{2\text{Gy}}(\alpha_X) < S_{X\%}^{9.2\text{Gy}}(\alpha_X)$  and  $S_{X\%}^{2\text{Gy}}(\beta_X) > S_{X\%}^{9.2\text{Gy}}(\beta_X)$ . In general,  $S_{X\%}^{PTV}(\beta_X) > S_{X\%}^{optic}(\beta_X)$ .

For the RMF model, a comparison of the reduced uncertainty analysis results (Table III) to the comprehensive uncertainty results in Table II shows that the comprehensive  $\sigma_{RWD_{X\%}}$  is larger than the reduced  $\sigma_{RWD_{X\%}}$ . The reduced uncertainty analysis can serve as a lower limit to the comprehensive uncertainty analysis.

### 4. DISCUSSION

#### 4.A. Implementation

We showed the feasibility of variance-based SA for RBE and RWD calculations in carbon ion therapy. Due to the Monte Carlo approach, an uncertainty as well as a SA was retrieved. This can be implemented for any relevant value used for treatment plan assessment (e.g., dose maps, DVHs, equivalent uniform dose, or RWD<sub>X%</sub>). The needed computation time and used memory are the biggest challenge for a convenient application in treatment planning. They can be reduced in several ways: reduced number of evaluated voxels or necessary runs and improved hardware and dedicated implementations.

Several improvements are hence conceivable: The SA can be only performed for a specified region of interest (e.g., in the PTV or a suitable extension of the PTV), reducing the

number of evaluated voxels. The use of low discrepancy quasi random numbers (e.g., Sobols' sequence<sup>36,40</sup>) for the Monte Carlo sampling potentially reduces the necessary  $n_{run}$ . The computational problem itself is parallelizable to a very high degree, meaning that doubled computational power runs twice as fast. In terms of hardware upgrades, also the use of dedicated high-end GPUs is promising, as the necessary operations are fairly simple, just need to be executed very often.

A future application of the presented approach is in the field of robust optimization. The Monte Carlo approach necessary for a variance-based SA provides an uncertainty of the examined result which can be used for the optimization. A first possible scenario how to use the SA for a new robust optimization procedure can be found in Table II: RWD prediction are far more sensitive to  $c_2$  for  $\alpha_x/\beta_x = 2$  Gy than it is the case for  $\alpha_x/\beta_x = 9.2$  Gy. This means that in the corresponding robust optimization, the main objective is a beam arrangement minimizing the impact of uncertainties in  $c_2$ . Similar conclusions can be drawn for different OARs, depending on their proximity to the PTV.

Besides this also an optimization of SA values themselves can be foreseen, e.g., optimizing a treatment plan in a way that for example an unavoidable uncertainty in the  $\alpha_x/\beta_x$  has a minimal  $S_k$  value, indicating that the treatment plan is the least vulnerable to it. Note that the computational performance of the SA needs to be improved for this kind of SA optimization, as it has to be executed several times in every optimization step.

A potential limitation that needs to be considered is the necessity to assign distributions to the input factors. Often

TABLE III. RWD<sub>2%</sub>, RWD<sub>50%</sub>, and RWD<sub>98%</sub> and their uncertainties ( $\sigma_{RWD_{X\%}}$ ). The result of the sensitivity analysis is reported for the PTV and left optic nerve. Two representative  $\alpha_x/\beta_x$  values were evaluated with the RMF model and the LEM1, respectively. The uncertainty and sensitivity analysis was done for RWD sampling input uncertainties for  $\alpha_x$ ,  $\beta_x$ , and  $d$ , whereas  $\alpha_p$  and  $\beta_p$  were kept constant. The corresponding RW-DVHs are displayed in Fig. S2.

$\alpha_x/\beta_x = 2$ Gy $RWD_{X\%} \pm \sigma_{RWD_{X\%}}$ [Gy (RBE)]	PTV, RMF			$\alpha_x/\beta_x = 9.2$ Gy $RWD_{X\%} \pm \sigma_{RWD_{X\%}}$ [Gy (RBE)]	PTV, RMF				
	$\alpha_x$	$\beta_x$	$d$		$\alpha_x$	$\beta_x$	$d$		
$RWD_{2\%} = 3.02 \pm 0.18$	$S_{2\%}$	0.18	0.41	0.42	$RWD_{2\%} = 3.03 \pm 0.24$	$S_{2\%}$	0.59	0.07	0.34
$RWD_{50\%} = 3.00 \pm 0.18$	$S_{50\%}$	0.17	0.40	0.43	$RWD_{50\%} = 3.00 \pm 0.24$	$S_{50\%}$	0.59	0.07	0.35
$RWD_{98\%} = 2.94 \pm 0.18$	$S_{98\%}$	0.18	0.40	0.42	$RWD_{98\%} = 2.94 \pm 0.23$	$S_{98\%}$	0.60	0.06	0.34
	PTV, LEM1				PTV, LEM1				
$RWD_{2\%} = 3.02 \pm 0.17$	$S_{2\%}$	0.20	0.47	0.34	$RWD_{2\%} = 3.02 \pm 0.23$	$S_{2\%}$	0.63	0.07	0.30
$RWD_{50\%} = 3.00 \pm 0.17$	$S_{50\%}$	0.20	0.46	0.34	$RWD_{50\%} = 3.00 \pm 0.23$	$S_{50\%}$	0.63	0.07	0.30
$RWD_{98\%} = 2.94 \pm 0.17$	$S_{98\%}$	0.21	0.46	0.34	$RWD_{98\%} = 2.94 \pm 0.23$	$S_{98\%}$	0.64	0.07	0.30
$\alpha_x/\beta_x = 2$ Gy $RWD_{X\%} \pm \sigma_{RWD_{X\%}}$ [Gy (RBE)]	Optic nerve, RMF			$\alpha_x/\beta_x = 9.2$ Gy $RWD_{X\%} \pm \sigma_{RWD_{X\%}}$ [Gy (RBE)]	Optic nerve, RMF				
	$\alpha_x$	$\beta_x$	$d$		$\alpha_x$	$\beta_x$	$d$		
$RWD_{2\%} = 2.19 \pm 0.13$	$S_{2\%}$	0.28	0.34	0.38	$RWD_{2\%} = 2.10 \pm 0.18$	$S_{2\%}$	0.67	0.04	0.29
$RWD_{50\%} = 1.20 \pm 0.08$	$S_{50\%}$	0.49	0.18	0.33	$RWD_{50\%} = 1.07 \pm 0.10$	$S_{50\%}$	0.75	0.01	0.24
$RWD_{98\%} = 0.18 \pm 0.02$	$S_{98\%}$	0.75	0.01	0.22	$RWD_{98\%} = 0.15 \pm 0.02$	$S_{98\%}$	0.78	0.00	0.19
	Optic nerve, LEM1				Optic nerve, LEM1				
$RWD_{2\%} = 2.25 \pm 0.13$	$S_{2\%}$	0.29	0.38	0.34	$RWD_{2\%} = 2.15 \pm 0.17$	$S_{2\%}$	0.69	0.04	0.27
$RWD_{50\%} = 1.31 \pm 0.08$	$S_{50\%}$	0.47	0.21	0.32	$RWD_{50\%} = 1.11 \pm 0.11$	$S_{50\%}$	0.75	0.01	0.23
$RWD_{98\%} = 0.28 \pm 0.03$	$S_{98\%}$	0.74	0.02	0.24	$RWD_{98\%} = 0.17 \pm 0.02$	$S_{98\%}$	0.79	0.00	0.20



exact values might not be available. Further studies are necessary to evaluate the impact of potentially errors in the assigned distributions. In a first simulation, all  $S_k$  and their relative trends stayed constant within 1% when changing input distributions from normal to uniform ( $\pm 45\%$  for  $\alpha_x/\beta_x$ ,  $\pm 15\%$  for  $d$ ,  $\pm 30\%$  for  $c_1$ , and  $c_2$ ). This is consistent with the results reported by Kamp et al.<sup>20</sup> and might indicate that the actual (unknown) shape of the uncertainty distributions is less relevant.

A main limitation of the used SA approach is that it is not suitable for correlated input uncertainties. The mathematical definition of  $S_k$  breaks down for correlated inputs. Even though the uncertainty analysis could still be carried out with correlated input sampling, the SA itself can only be executed with uncorrelated and independent input parameter sampling.

A more general limitation of the presented SA approach is that it relies in the LQ model approximations for RBE predictions. The simulated RWD variations are only derived in the framework of this commonly used model. In this respect also systematic differences in the *in vivo* and *in vitro* data are not explicitly modeled and were hence not included in the SA. Nonetheless, the SA adds and analyses systematic uncertainties to RBE and RWD predictions. The mentioned shortcomings do also adhere to clinical treatment plans which are currently based on RBE predictions in the LQ framework.

Besides the uncertainties in the RBE modeling process, carbon ion treatment planning has to take uncertainties in patient setup and range into account. In contrast to the biological uncertainties, patient setup and range uncertainties can be accounted for by margin concepts in clinical practice. In the presented first patient case potential range and setup uncertainties are summarized in a 5% uncertainty in the physical dose. The variance-based sensitivity approach is in general suitable to evaluate more complex scenarios, e.g., explicitly including these mentioned range and setup uncertainties.

#### 4.B. Sensitivity maps for RMF model predictions

The result that  $S_{z_p}(\alpha_x) = S_{z_p}(c_1)$  and  $S_{z_p}(\beta_x) = S_{z_p}(c_2)$  is direct consequence of the assigned relative standard deviations and the calculation of  $\alpha_p$  in Eq. (2). For example, considering the first term of the sum, a relative change of +10% in either  $\alpha_x$  or  $c_1$  results in the same multiplication by 1.1. The spatial distribution in the sensitivity maps is different for panels g and j (or h and k) of Fig. 1.  $\alpha_p$  is more dependent on changes in  $\alpha_x$  and  $c_1$  in voxels with lower LET. For high LET regions, uncertainties in  $\beta_x$  and  $c_2$  have a higher impact. This can be explained by the strong dependence of  $c_2$  on LET (compare to panel b and i of Fig. 2). The spatially constant  $S_{\beta_p}(\beta_x)$  and  $S_{\beta_p}(c_1)$  result from Eq. (3) and the relative definition of the uncertainties.

The comprehensive SA for  $RWD_{RMF}(\alpha_x/\beta_x, c_1, c_2, d)$  are shown in Fig. 2. The high values in  $S_{RWD}(\alpha_x/\beta_x)$  and  $S_{RWD}(c_2)$  and the corresponding low  $S_{RWD}(c_1)$  are in regions of high LET. A strong LET dependency can already be seen for the unchanged  $c_1$  and  $c_2$  (panels h

and i). This dependency on LET in combination with the changing sensitivity is of interest, especially if an organ at risk (OAR) lies in this region. In the presented patient case the left optic nerve, marked in green in the panels, is located in the region of high LET. The uncertainty in  $\alpha_x/\beta_x$  is more crucial for this OAR than for the PTV. The part of the biological modeling represented by  $c_1$  on the other hand shows the highest impact on the uncertainty of RWD in the PTV but only a small impact on the uncertainty of RWD in the left optic nerve.

The sensitivity values in the fragmentation tail, distant from the PTV can be verified analytically. For small  $d$ , RBE reduces to  $RBE_\alpha \approx \alpha_p/\alpha_x = c_1 + c_2 \cdot \beta_x/\alpha_x$ . The contribution of  $c_2$  can be neglected in the fragmentation tail. Hence,  $S_{RBE}(c_1) > 0.9$  (compare to panel k of Fig. S1. RWD in this region is calculated by multiplying  $RBE_\alpha$  with  $d$ , resulting in  $S_{RWD}(c_1) \approx 0.68$  and  $S_{RWD}(d) \approx 0.24$  (panel f and k in Fig. 2), showing that  $c_1$  and  $d$  are the only influencing factors ( $S_{RWD}(c_2) \approx 0.02$  and  $S_{RWD}(\alpha_x/\beta_x) \approx 0.02$ ).

#### 4.C. Structure-based sensitivity analysis for RMF model predictions

The result of the optimized RWD distribution is very similar for the two considered  $\alpha_x/\beta_x$ . A comparison of the different  $\alpha_x/\beta_x$  revealed different sensitivity values for  $c_1$  and  $c_2$ . This shows the potential of a variance-based SA. The most influential input uncertainties are revealed, meaning that a robust optimization scenario could optimize the influence of different model parameters for different  $\alpha_x/\beta_x$ .

The SA of  $RWD_{X\%}$  in Table II for  $\alpha_x/\beta_x = 2$  Gy combine the 3D results in the sensitivity maps in Fig. 2. The trends observed in the axial slices can also be seen here:  $S(c_1)$  are higher in low-medium LET region of the PTV and  $S(\alpha_x/\beta_x)$  and  $S(c_2)$  are higher in high LET regions (optic nerve).

$S_{X\%}(\alpha_x/\beta_x) + S_{X\%}(d) + S_{X\%}(c_1) + S_{X\%}(c_2) \approx 1$  for all shown structures and cell lines. This means there are no relevant cross terms and hence interplay effects between the four inputs.

Removing the 5% uncertainty in  $d$  only has limited effect on  $\sigma_{RWD_{X\%}}$ . This means that even if the physical dose is delivered exactly as planned, a large uncertainty remains which is associated with the uncertainty in the biological modeling. This effect becomes more dramatic considering that a  $\sigma_d$  was set rather large, whereas an uncertainty of 10% in the biological parameters ( $\alpha_x$ ,  $\beta_x$ ,  $c_1$ , and  $c_2$ ) can be considered rather small.

#### 4.D. Reduced SA for different biological models

The reduced uncertainty analysis is a computational inexpensive lower estimate for the comprehensive uncertainty. This lower uncertainty estimate might be a good candidate for a SA triggered robust optimization, which can in a first step be based “only” a reduced but fast SA. Further studies are needed to show how to use this in, e.g., multistage robust

optimization, consisting of a first stage based on the reduced SA and a second on a comprehensive SA.

Compared to the decoupled implementation of the RMF model,<sup>33</sup> the calculations and simulations needed for the LEM1 are in general more complex<sup>18</sup> and a comprehensive SA, including all LEM1 parameters was not yet computationally feasible. The reduced SA presented here shows a first step in the direction of a comprehensive SA for different biological models. In such a comprehensive SA, it can be expected that for different biological models different parameters have different effect on  $\sigma_{RWD_{X\%}}$ . An example for this is the finding that  $S_{X\%}^{RMF}(d) > S_{X\%}^{LEM1}(d)$  which can be directly observed from Eq. (1) where  $d$  is multiplied with  $\beta_p$ . Considering that LEM1 predictions of  $\beta_p$  decrease with increasing LET<sup>37</sup>, whereas RMF model predictions increase with increasing LET<sup>28,32</sup> the influence of uncertainty in  $d$  on  $\sigma_{RWD_{X\%}}$  will be higher for RMF model predictions.

## 5. CONCLUSION

A variance-based SA including an uncertainty analysis for carbon ion treatment plans is feasible. The proposed approach provides new insight into robustness of treatment plans. Both evaluated biological models showed similar uncertainty bands in the reduced SA when  $\alpha_X$ ,  $\beta_X$ , and  $d$  were subject to uncertainties. In general, it can be concluded that input parameter sensitivity varies depending on the spatial location in the treatment plan. Due to the high computational effort resulting from the necessary Monte Carlo sampling, a comprehensive SA is currently only possible using the RMF model. The described SA analysis provides two new sources of information: Monte Carlo-derived uncertainty and variance-based sensitivity information, breaking down the uncertainty to the different input parameters. The main potential applications are seen in manual treatment plan evaluation and in biologically robust treatment plan optimization.

## ACKNOWLEDGMENT

The authors thank G. Cabal, A. Mairani, and K. Parodi for providing the used fragment spectra for carbon ion beams and D.J. Carlson for his support with the repair-misrepair-fixation model. This work was partially supported by a fellowship from the German Academic Exchange Service (DAAD), the DFG grant WI 3745/1-1 and DFG cluster of excellence: Munich-Centre for Advanced Photonics.

## CONFLICTS OF INTEREST

The authors have no conflicts to disclose.

<sup>a)</sup> Author to whom correspondence should be addressed. Electronic mail: florian.kamp@mytum.de.

## REFERENCES

- Kellerer AM, Rossi HH. A generalized formulation of dual radiation action. *Radiat Res.* 1978;75:471–488.
- Liebl J, Paganetti H, Zhu M, Winey BA. The influence of patient positioning uncertainties in proton radiotherapy on proton range and dose distributions. *Med Phys.* 2014;41:91711.
- Lomax AJ. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties. *Phys Med Biol.* 2008;53:1027–1042.
- Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Phys Med Biol.* 2012;57:R99–R117.
- Enghardt W, Crespo P, Fiedler F, et al. Charged hadron tumour therapy monitoring by means of PET. *Nucl Instrum Methods Phys Res, Sect A.* 2004;525:284–288.
- Bauer J, Unholtz D, Sommerer F, et al. Implementation and initial clinical experience of offline PET/CT-based verification of scanned carbon ion treatment. *Radiother Oncol.* 2013;107:218–226.
- Pinto M, Bajard M, Brons S, et al. Absolute prompt-gamma yield measurements for ion beam therapy monitoring. *Phys Med Biol.* 2015a;60:565–594.
- Pinto M, de Rydt M, Dauvergne D, et al. Technical note: experimental carbon ion range verification in inhomogeneous phantoms using prompt gammas. *Med Phys.* 2015b;42:2342–2346.
- Pönisch F, Parodi K, Hasch BG, Enghardt W. The modelling of positron emitter production and PET imaging during carbon ion therapy. *Phys Med Biol.* 2004;49:5217–5232.
- Bangert M, Hennig P, Oelfke U. Analytical probabilistic modeling for radiation therapy treatment planning. *Phys Med Biol.* 2013;58:5401–5419.
- Chen W, Unkelbach J, Trofimov A, et al. Including robustness in multi-criteria optimization for intensity-modulated proton therapy. *Phys Med Biol.* 2012;57:591–608.
- Pflugfelder D, Wilkens JJ, Oelfke U. Worst case optimization: a method to account for uncertainties in the optimization of intensity modulated proton therapy. *Phys Med Biol.* 2008;53:1689–1700.
- Unkelbach J, Chan TCY, Bortfeld T. Accounting for range uncertainties in the optimization of intensity modulated proton therapy. *Phys Med Biol.* 2007;52:2755–2773.
- Fredriksson A, Forsgren A, Hårdemark B. Maximizing the probability of satisfying the clinical goals in radiation therapy treatment planning under setup uncertainty. *Med Phys.* 2015;42:3992–3999.
- Liu W, Zhang X, Li Y, Mohan R. Robust optimization of intensity modulated proton therapy. *Med Phys.* 2012;39:1079–1091.
- Fredriksson A. A characterization of robust radiation therapy treatment planning methods—from expected value to worst case optimization. *Med Phys.* 2012;39:5169–5181.
- Böhlen TT, Brons S, Dosanjh M, et al. Investigating the robustness of ion beam therapy treatment plans to uncertainties in biological treatment parameters. *Phys Med Biol.* 2012;57:7983–8004.
- Friedrich T, Grün R, Scholz U, Elsässer T, Durante M, Scholz M. Sensitivity analysis of the relative biological effectiveness predicted by the local effect model. *Phys Med Biol.* 2013a;58:6827–6849.
- Friedrich T, Weyrather W, Elsässer T, Durante M, Scholz M. Accuracy of RBE: experimental and theoretical considerations. *Radiat Environ Biophys.* 2010;49:345–349.
- Kamp F, Brüningk SC, Cabal G, Mairani A, Parodi K, Wilkens JJ. Variance-based sensitivity analysis of biological uncertainties in carbon ion therapy. *Phys Med.* 2014;30:583–587.
- Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research. *Med Phys.* 2003;30:979–985.
- Schell S, Wilkens JJ. Advanced treatment planning methods for efficient radiation therapy with laser accelerated proton and ion beams. *Med Phys.* 2010;37:5330–5340.
- Brüningk SC, Kamp F, Wilkens JJ. EUD-based biological optimization for carbon ion therapy. *Med Phys.* 2015;42:6248.
- Kamp F, Cabal G, Mairani A, Parodi K, Wilkens JJ, Carlson DJ. Fast biological modeling for voxel-based heavy ion treatment planning using the mechanistic repair-misrepair-fixation model and nuclear fragment spectra. *Int J Radiat Oncol Biol Phys.* 2015;93:557–568.
- Wilkens JJ, Oelfke U. Fast multifield optimization of the biological effect in ion therapy. *Phys Med Biol.* 2006;51:3127–3140.



26. Parodi K, Mairani A, Brons S, et al. Monte Carlo simulations to support start-up and treatment planning of scanned proton and carbon ion therapy at a synchrotron-based facility. *Phys Med Biol.* 2012;57:3759–3784.
27. Scholz M, Kellerer AM, Kraft-Weyrather W, Kraft G. Computation of cell survival in heavy ion beams for therapy. *Radiat Environ Biophys.* 1997;36:59–66.
28. Carlson DJ, Stewart RD, Semenenko VA, Sandison GA. Combined use of Monte Carlo DNA damage simulations and deterministic repair models to examine putative mechanisms of cell killing. *Radiat Res.* 2008;169:447–459.
29. Semenenko VA, Stewart RD. A fast Monte Carlo algorithm to simulate the spectrum of DNA damages formed by ionizing radiation. *Radiat Res.* 2004;161:451–457.
30. Semenenko VA, Stewart RD. Fast Monte Carlo simulation of DNA damage formed by electrons and light ions. *Phys Med Biol.* 2006;51:1693–1706.
31. Stewart RD, Yu VK, Georgakilas AG, Koumenis C, Park JH, Carlson DJ. Effects of radiation quality and oxygen on clustered DNA lesions and cell death. *Radiat Res.* 2011;176:587–602.
32. Frese MC, Yu VK, Stewart RD, Carlson DJ. A mechanism-based approach to predict the relative biological effectiveness of protons and carbon ions in radiation therapy. *Int J Radiat Oncol Biol Phys.* 2012;83:442–450.
33. Kamp F, Carlson DJ, Wilkens JJ. Rapid implementation of the repair-misrepair-fixation (RMF) model facilitating online adaption of radiosensitivity parameters in ion therapy. *Phys Med Biol.* 2017;62:N285–N296.
34. Mairani A, Dokic I, Magro G, et al. Biologically optimized helium ion plans: calculation approach and its in vitro validation. *Phys Med Biol.* 2016;61:4283–4299.
35. Weyrather WK, Ritter S, Scholz M, Kraft G. RBE for carbon track-segment irradiation in cell lines of differing repair capacity. *Int J Radiat Biol.* 1999;75:1357–1364.
36. Saltelli A, Ratto M, Andres T, et al. *Global Sensitivity Analysis: The Primer.* Chichester, England/Hoboken, NJ: John Wiley; 2008.
37. Friedrich T, Scholz U, Elsässer T, Durante M, Scholz M. Systematic analysis of RBE and related quantities using a database of cell survival experiments with ion beam irradiation. *J Radiat Res.* 2013b;54:494–514.
38. Schulz-Ertner D, Karger CP, Feuerhake A, et al. Effectiveness of carbon ion radiotherapy in the treatment of skull-base chordomas. *Int J Radiat Oncol Biol Phys.* 2007;68:449–457.
39. Furusawa Y, Fukutsu K, Aoki M, et al. Inactivation of aerobic and hypoxic cells from three different cell lines by accelerated 3 He-, 12 C- and 20 Ne-ion beams. *Radiat Res.* 2000;154:485–496.
40. Sobol' IM. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput Math Math Phys.* 1967;7:86–112.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Fig. S1.** Sensitivity maps for  $RBE_{RMF}(\alpha_x/\beta_x, c_1, c_2, d)$ . The RBE (panel a) is displayed together with its uncertainty (panel d). The LET (panel b) and the beam geometry (panel e) are shown together with the four unchanged input parameter maps (panels c, g–i). The four sensitivity maps are shown in Panels (f), (j)–(l). The planning target (PTV) volume is marked in red, the left optical nerve in green together with the left eyeball (orange), and the left lens (brown). The underlying treatment plan optimized for a RWD of 3 Gy(RBE) can be seen in Fig. 2.

**Fig. S2.** RW-DVH for the PTV and the left optic nerve for two representative  $\alpha_x/\beta_x$ : 2 Gy (panels a and c) and 9.2 Gy (panels b and d). Panels (a) and (b) display the result using the RMF model — panels (c) and (d) the result for LEM1 predictions. The black RW-DVH lines denote the result of the unchanged optimization result. The uncertainties in the RW-DVH are indicated as  $\pm\sigma$  (68.3% probability to lie inside the dark area) and  $\pm 2\sigma$  (95.4% probability to lie inside the light and dark area). The uncertainty and sensitivity analysis was done for RWD being a function of  $\alpha_x$ ,  $\beta_x$ , and  $d$ , whereas  $\alpha_p$  and  $\beta_p$  were kept constant. Table III summarizes the main sensitivity values corresponding to this figure.



Dose guided positioning

## Multi-criterial patient positioning based on dose recalculation on scatter-corrected CBCT images



Jan Hofmaier<sup>a,c,\*</sup>, Jonas Haehnle<sup>b</sup>, Christopher Kurz<sup>a,c</sup>, Guillaume Landry<sup>c</sup>, Cornelius Maihoefer<sup>a,g</sup>, Lars Schüttrumpf<sup>a,g</sup>, Philipp Süß<sup>b</sup>, Katrin Teichert<sup>b</sup>, Matthias Söhn<sup>a</sup>, Nadine Spahr<sup>d</sup>, Christoph Brachmann<sup>e</sup>, Florian Weiler<sup>e</sup>, Christian Thieke<sup>a</sup>, Karl-Heinz Küfer<sup>b</sup>, Claus Belka<sup>a,f,g</sup>, Katia Parodi<sup>c</sup>, Florian Kamp<sup>a</sup>

<sup>a</sup> Department of Radiation Oncology, University Hospital, LMU Munich, Munich; <sup>b</sup> Fraunhofer Institute for Industrial Mathematics ITWM, Kaiserslautern; <sup>c</sup> Department of Medical Physics, Faculty of Physics, Ludwig-Maximilians-Universität München, Munich; <sup>d</sup> Fraunhofer Institute for Medical Image Computing MEVIS, Lübeck; <sup>e</sup> Fraunhofer Institute for Medical Image Computing MEVIS, Bremen; <sup>f</sup> German Cancer Consortium (DKTK), Munich; and <sup>g</sup> Clinical Cooperation Group Personalized Radiotherapy in Head and Neck Cancer, Helmholtz Zentrum München, Munich, Germany

### ARTICLE INFO

#### Article history:

Received 30 May 2017

Received in revised form 18 September 2017

Accepted 19 September 2017

Available online 12 October 2017

This work has been partially presented at the ESTRO 36 meeting in Vienna, May 2017.

#### Keywords:

Isocenter correction

Scatter correction

CBCT

Dose guided positioning

### ABSTRACT

**Background and purpose:** Our aim was to evaluate the feasibility and potential advantages of dose guided patient positioning based on dose recalculation on scatter corrected cone beam computed tomography (CBCT) image data.

**Material and methods:** A scatter correction approach has been employed to enable dose calculations on CBCT images. A recently proposed tool for interactive multicriterial dose-guided patient positioning which uses interpolation between pre-calculated sample doses has been utilized. The workflow was retrospectively evaluated for two head and neck patients with a total of 39 CBCTs. Dose–volume histogram (DVH) parameters were compared to rigid image registration based isocenter corrections (clinical scenario).

**Results:** The accuracy of the dose interpolation was found sufficient, facilitating the implementation of dose guided patient positioning. Compared to the clinical scenario, the mean dose to the parotid glands could be improved for 2 out of 5 fractions for the first patient while other parameters were preserved. For the second patient, the mean coverage over all fractions of the high dose PTV could be improved by 4%. For this patient, coverage improvements had to be traded against organ at risk (OAR) doses within their clinical tolerance limits.

**Conclusions:** Dose guided patient positioning using in-room CBCT data is feasible and offers increased control over target coverage and doses to OARs.

© 2017 Elsevier B.V. All rights reserved. Radiotherapy and Oncology 125 (2017) 464–469

In modern intensity-modulated radiotherapy (IMRT), a treatment plan is optimized in order to cover the target volume with the prescribed dose while optimally sparing adjacent healthy structures. The optimization takes the individual patient geometry into account, which is imaged before the beginning of the treatment course by a computed tomography (CT) scan. To ensure that the delivered dose is in agreement with the calculated dose in the treatment planning system, the patient needs to be positioned and aligned to the treatment unit in a reproducible and accurate way. Typically, this is performed using in-room imaging equipment such as cone-beam computed tomography (CBCT) [1]. A rigid registration with 6 or 3 degrees of freedom between planning CT (pCT)

and CBCT is performed, and the patient table is moved accordingly to account for set-up errors [2]. The rigid registration is typically based on anatomical landmarks close to the tumor (e.g. bones) or implanted fiducial markers [3]. However, during the course of treatment, considerable non-rigid changes may occur, such as tumor shrinkage or weight-loss [4,5]. In this case, the actual dose distribution might differ substantially from the one calculated on the pCT [6], and determining the clinically favorable rigid isocenter correction is not obvious. This could possibly result in both a clinically relevant underdosage of the tumor and an overdosage of relevant organs at risk (OAR), thus risking reduced tumor control and increasing toxicity. Due to the high scatter contribution, CT numbers of CBCTs are not sufficiently accurate for dose calculation. However, lately there has been a lot of progress in intensity correction strategies for CBCT to enable photon and proton dose calculation on these images [7–9], and it seems therefore reasonable to

\* Corresponding author at: Klinikum der Universität München, Klinik und Poliklinik für Strahlentherapie und Radioonkologie, Marchioninistraße 15, D-81377 München, Germany.

E-mail address: [jan.hofmaier@med.uni-muenchen.de](mailto:jan.hofmaier@med.uni-muenchen.de) (J. Hofmaier).

base the isocenter correction on a dose recalculation on the image of the day rather than on the clinically used rigid image registration between pCT and CBCT. This has first been suggested for proton therapy [6,10]. Recently, an interactive, multicriterial approach to dose-guided patient positioning has been presented and evaluated for re-planning CTs of head and neck (H&N) and prostate patients treated with photon IMRT [11]. In the present work, we evaluate the potential of the method using dose recalculation on scatter corrected CBCT (scCBCT) images.

## Methods

Fig. 1 gives an overview of the overall workflow. The three main features are a scatter correction method for CBCTs, a propagation of contours from the pCT to the scCBCT and an interactive, pareto optimal dose-based isocenter correction, balancing dose–volume histogram (DVH) objectives. The workflow has been evaluated for two H&N patient cases with a total of 39 CBCTs.

### Imaging and rigid registration based isocenter correction

The pCTs were acquired with a Toshiba Aquilion CT scanner. The CBCT images were acquired with the integrated scanner in the Elekta Axesse system. The clinical rigid registration between CBCT and pCT was applied using a bony match in the Elekta XVI software. Isocenter position corrections were restricted to 3 degrees of freedom, i.e. only translational corrections.

### Scatter correction and CBCT contouring

After the CBCT acquisition, a scatter correction of the CBCT is performed. Our implementation of scatter correction closely followed [8,9]. Virtual CTs (vCTs) obtained by deformable image registration (DIR) are used as a prior to estimate the scatter contribution in the CBCT projections.

The workflow of the scatter correction is as follows (as also highlighted in Fig. 1):

1. DIR of the pCT to the CBCT to obtain a vCT “of the day”. The vector field is also used to propagate the contours from the pCT to the vCT.
2. Forward projection of the vCT according to the cone beam geometry.
3. Application of a correction factor (CF) to the CBCT projections to match the intensities to the vCT forward projections. Following [8], we used  $CF = 25.6$ .
4. Estimation of scatter by subtraction of the vCT forward projections from the CBCT projections and a smoothing operation (a 2D median filter with 25-by-25 pixels width, followed by a Gaussian filter of 1.5 pixels standard deviation).
5. Subtraction of scatter from the CBCT projections.
6. Reconstruction of the scCBCT.

The advantage of this approach over using the vCT itself for dose calculation is that it is insensitive to small errors in the DIR and contrast of the CBCT is not altered by the scatter-correction [7,8].

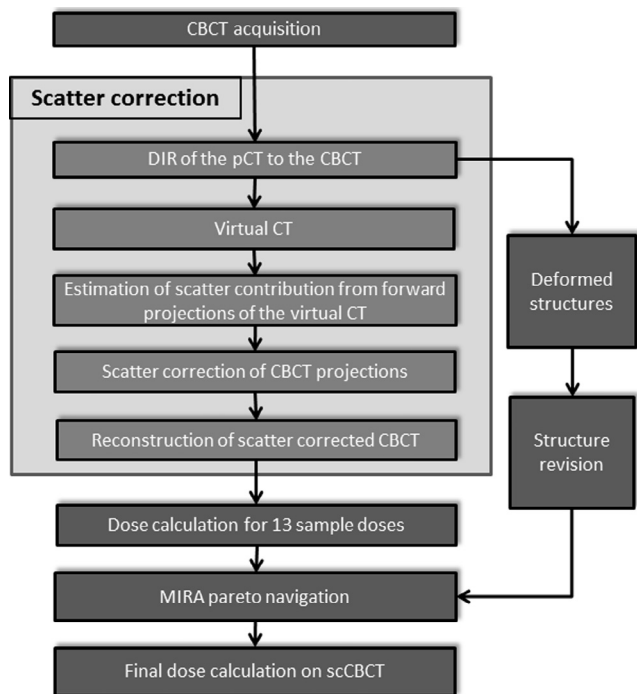
In our workflow, a variational approach is used for DIR and aims for image similarity and deformation regularity. Image similarity is measured by Normalized Gradient Fields (NGF) [12] and deformation regularity is modeled by curvature regularization [13]. The resulting optimization problem is solved in a discretize-then-optimize scheme using a quasi-Newton L-BFGS optimizer.

Due to the limited field of view (FOV) of the CBCT and hence the scCBCT, the corresponding regions of the pCT outside the FOV were stitched to the scCBCT. This was necessary to account for beam attenuation when irradiating through the shoulders of the patients, which were not covered by the CBCT FOV. A comparison of the different types of CTs can be found in Supplemental Fig. 1.

The propagated contours of organs at risk (OARs) were adapted by a trained clinical expert. The PTV contours were adapted by senior physicians.

### Dose calculation and interpolation

The dose distribution of the original clinical plan on the scCBCT was calculated using a Monte Carlo (MC) algorithm on a  $3 \times 3 \times 3$  mm<sup>3</sup> dose grid. The used dose engine was MCverify v2.44, a scriptable research version of the same algorithm used in Elekta Monaco 5.1. Clinical treatment plans of the evaluated patients were delivered with an Elekta Axesse LINAC. For every scCBCT, the dose was calculated for 13 sample isocenter positions: the central position, based on a rigid alignment of the CBCT to the pCT, and shifts of  $\pm 3$  mm and  $\pm 6$  mm along every axis. For each fraction, the clinically delivered plan was used for the dose calculations on the scCBCTs (also in case of a clinical re-planning). As described in more detail in [11], linear combinations of the sample dose distributions are used to estimate dose distributions in a continuous space of possible isocenter shifts. To avoid errors in our evaluation of DVH parameters, which might be introduced by inaccuracies in the dose interpolation, also a final dose calculation was performed once a satisfying isocenter correction was found. Reported DVH parameters were therefore always determined from a dose calculation for the final isocenter position without interpolation involved. Interpolated doses were only used during the multicriterial optimization described below to find that isocenter position. The interpolation accuracy was evaluated by interpolating dose cubes for random 3D isocenter shifts within the accessible range of the interpolation and comparing them to the respective forward calculation using a 2% dose difference criterion and a 2%/2 mm gamma criterion. 100 random shifts for each scCBCT of patient 1 and 30 random shifts for each scCBCT of patient 2 were evaluated.



**Fig. 1.** Overview of the applied dose guided positioning workflow. The deformable image registration is performed both for the scatter correction and to obtain deformed contours, which are used as a starting point for the contouring of the CBCTs. To avoid any interpolation errors in the evaluation, a final dose calculation was added once a satisfying isocenter correction was found.

### Multicriterial interactive radiotherapy assistant

The multicriterial interactive radiotherapy assistant (MIRA) is a research treatment planning system for IMRT. The interface gives the user the possibility to browse pareto-optimal plan candidates interactively [14–16]. The recently implemented multicriterial isocenter optimization in MIRA has been described in [11]. Pre-calculated sample doses for a set of isocenter shifts and interpolation between them are used to enable interactive navigation in real-time. The multicriterial approach is DVH based, meaning that the user doesn't manipulate the isocenter directly. Instead, the tool offers sliders corresponding to DVH deviation cost functions for the structures of interest. These cost functions compare a DVH on the pCT with the respective DVH on the scCBCT for a possible isocenter position. For OARs, they are defined as follows:

$$q_{OAR}(d^{c,scCBCT}, d^{pCT}) = \sqrt{\sum_{i=0}^{100} \max(0, D_i^{c,scCBCT} - D_i^{pCT})^2}$$

where  $c$  denotes a possible isocenter shift,  $d$  denotes the respective dose distributions on pCT and scCBCT and  $D_{0...100}$  are the dose quantiles of the corresponding cumulative DVH. It is a one-sided cost function, which penalizes any of the  $D_i^{scCBCT}$  being larger than the respective  $D_i^{pCT}$ .

For target structures the cost function is defined analogously:

$$q_{Target}(d^{c,scCBCT}, d^{pCT}) = \sqrt{\sum_{i=0}^{49} \max(0, D_i^{c,scCBCT} - D_i^{pCT})^2 + \sum_{i=50}^{100} \max(0, D_i^{pCT} - D_i^{c,scCBCT})^2}$$

This cost function penalizes any increase of the dose quantiles  $D_{0...49}$  and any decrease of the dose quantiles  $D_{50...100}$ , therefore it maximizes target coverage while still minimizing overdosage.

The multicriterial optimization restricts the accessible isocenters to the subset of pareto-optimal solutions – therefore, no DVH objective can be improved without worsening another. When the user manipulates the sliders, he immediately gets the result of this trade-off. The tool is up to now restricted to translational shifts in 3 dimensions and does not yet support rotational adjustments.

### Application to patient data

Clinical datasets of two patients who had received curative-intended radiotherapy for head and neck cancer have been retrospectively evaluated. Their characteristics are shown in Table 1. Patients were positioned in supine position and immobilized with a thermoplastic mask. The clinical step-and-shoot IMRT (ssIMRT) treatment plan for patient 1 was generated with Hyperion V2.42, a research version of the treatment planning system (TPS) Elekta Monaco. For validation, the plan was re-calculated in the TPS Oncentra Masterplan. The clinical VMAT plans for patient 2 were optimized in Monaco itself. Both patients were treated with a simultaneous integrated boost (SIB) concept. All calculated values for particular fractions were scaled to the total plan dose. Our evaluation focused on the trade-off between PTV coverage and dose to the parotid glands, which is of interest due to its association with xerostomia [17]. For patient 2, the left parotid gland was inside of the target volume, therefore, only the parotid gland on the right hand side was considered. For the other OARs, it was always

assured that the clinical constraints in Table 2 were met for the final dose guided isocenter shift.

## Results

### Interpolation accuracy

The median pass-rate for the 2% dose difference criterion over all 1520 calculated shifts was 92.8% (range 86.8–100.0%). Failing points were located predominantly at the patient surface, where the interpolation cannot be accurate. For a 2%/2 mm gamma criterion, median pass-rate was 99.0% (range 96.6–100.0%).

### DVH parameters

Table 3 shows mean values of relevant DVH parameters for both patients. For patient 1 the mean dose to both parotid glands could be improved compared to the clinical, rigid registration based shifts (31.0 to 30.2 Gy and 26.4 to 25.3 Gy, respectively). Other parameters were unchanged, except for a very small decrease in target coverage for the boost PTVs. Fig. 2 shows DVHs for one exemplary fraction of this patient. In this particular fraction, the mean dose to the left parotid gland improved from 30.6 to 26.1 Gy, the mean dose to the right parotid showed a small increase from 30.2 to 30.6 Gy. The coverage of the high dose PTV with the 95% isodose was slightly decreased. Isodose lines of a representative dose distribution for this patient can be found in Supplemental Fig. 2.

Fig. 3 shows the change of DVH parameters for patient 2 over time. For this patient, the mean coverage of the high dose PTV with the 95% isodose improved from 77% to 81% using dose guided positioning. The dose to the spared right parotid gland remained stable for this patient (mean value over all fractions of  $D_{mean}$  was 26.1 Gy both for the rigid registration based and dose guided isocenter corrections). A decrease in coverage occurs both for clinical as well as dose guided shifts around fraction 7, when large anatomical changes occurred due to necrotic degradation of the tumor (this is also visible in the plot of PTV size, Fig. 3f). This patient underwent offline-re-planning during his treatment course. The coverage is restored in fraction 18, when the original plan is replaced by the new one. After re-planning, the dose guided shifts perform better in preserving the coverage than the rigid image registration based shifts. However, improvements in coverage had to be traded against additional dose to the spinal cord (Fig. 2d and Table 3: the mean  $D_{2\%}$  increased from 37.4 to 38.3) and the brain stem (Table 3: mean  $D_{2\%}$  increased from 40.1 to 41.2).

The average Euclidean distance between the clinical and the optimized isocenter corrections for all available fractions was 1.8 mm (range 0.8–3.2 mm) for patient 1 and 2.0 mm (range 0.0–3.5 mm) for patient 2. A detailed plot of the difference between rigid registration based and dose guided isocenters can be found in Supplemental Fig. 3.

## Discussion

A recently proposed tool for dose guided patient positioning has been evaluated with in-room CBCT imaging data for the first time. The dose interpolation necessary for a fast multicriterial optimiza-

**Table 1**  
Evaluated patient cases.

No.	Prescription	Site of primary tumor	Technique	Sparing of parotid glands	Fractions with CBCT	Clinical re-planning	Primary tumor PTV volume change
1	66 Gy/60 Gy/54 Gy	Oral cavity	ssIMRT	bilateral	5/30	No	Stable (+4.8%)
2	70 Gy/56 Gy	Left cheek	VMAT	Only right side	34/35	Yes, after 17 fractions	Pronounced (–54.8%)



**Table 2**

OAR constraints that were always respected in the final dose isocenter shift.

Structure	Constraint
Spinal cord	$D_{2\%} \leq 45$ Gy
Optic chiasm	$D_{2\%} \leq 54$ Gy
Optic nerve	$D_{2\%} \leq 54$ Gy
Brain stem	$D_{2\%} \leq 54$ Gy

tion based on 13 sample isocenter positions was found accurate. In principle, the accuracy of the interpolation could be further improved with the inclusion of further sample points. A final forward calculation of the plan with the chosen isocenter shift was added for verification. A scatter correction approach was used to facilitate dose calculation on CBCTs. This approach was up to now only used and validated for proton dose calculation. A proton range agreement of less than 2 mm for more than 80% of the BEV profiles and a gamma passrate of more than 96% using a 2%/2 mm criterion for the proton dose distribution was reported comparing scCBCT and conventional CT [9]. Since photons are less prone to CT number inaccuracies than protons, it is safe to use scCBCTs for photon dose calculations. The validation in [9] was performed using the Elekta XVI system, which was also used in our study. Applying the scatter correction approach using other CBCT scanners might require a similar initial validation first. Also in [9], it was shown that the scatter correction approach can overcome small inaccuracies in the DIR, due to the smoothing function [8] applied to the scatter map. Due to the limited FOV of the scCBCT, regions of the pCT outside of the scCBCT FOV were stitched to the scCBCT to account for beam attenuation in these regions. The dosimetric uncertainty introduced by this approach is expected to be minimal, since head and neck are always fully covered by the FOV and in this region only the treatment table and patient immobilization devices are added, which are not subject to changes. In addition, the shoulders are added below the neck. Due to the treatment being coplanar, the uncertainty which might be introduced by an unprecise stitching affects only the lowest slices of the low dose PTV (lymphatic drainage area).

For patient 1, the mean dose to the parotid glands was improved compared to the clinical standard rigid anatomy based isocenter corrections, while other DVH parameters could be kept stable, except for a very small reduction in target coverage for the boost PTVs. For patient 2, coverage improvements had to be traded against OAR doses. For the affected OARs, spinal cord and brain stem, DVH parameters could be kept well below clinical constraints. The large maximum doses for the high dose PTV during the first days of treatment (Fig. 3e) are explained by the fact that this tumor was exulcerated and its volume had still increased by 8% between pCT (day 0 in Fig. 3f) and the first treatment fraction, which was delivered 7 days later. In order to improve the target

coverage at the surface of the tumor, the clinical treatment plan had high fluence tangentially to the tumor surface. The “auto flash margin” feature in Monaco, which opens the fields at the PTV borders when these are located at the patient surface, had then extended the high fluence regions beyond the patient surface. The tumor had grown into these fluence regions. In this particular case, this rather non-robust plan had been accepted, since the dose elevation only affected the exulcerated tumor and no healthy tissue was at risk. After the tumor volume decreased because parts of the tumor disappeared due to necrotic degradation around fraction 7, the maximum dose decreased to normal and the coverage of the PTV70 deteriorated. This also shows the limitations of this workflow: In the presence of such major anatomical changes, both rigid registration based and dose guided approach fail to restore an acceptable coverage of the target. At this time, clearly an adaptive online re-planning would have been beneficial. However, after the offline re-planning on day 18, the dose guided positioning approach performs considerably better than the rigid registration based approach in preserving target coverage in the presence of the smaller volume changes occurring during the rest of the treatment course.

In the presented implementation the isocenter corrections are limited to 3 translational degrees of freedom (d.o.f). In principle, the approach might be extended to 6 d.o.f. in future implementations, allowing also for rotational adjustments.

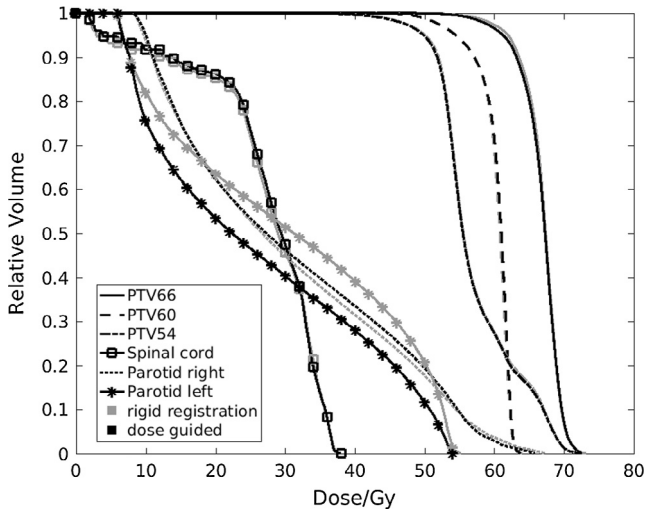
A drawback of the presented approach compared to the rigid registration based alignment is the need for contours on the daily image. In the presented workflow, contours are propagated from the pCT to the scCBCT using DIR. However, these contours need to be revised by physicians before a clinical decision can be made using them. In a future online implementation of the workflow, this will probably be the most time-consuming step. Besides the time needed for the contour adaption and their review also intra and inter observer variability remains a challenge for all studies and clinical workflows that require (re-) contouring. Other steps of the workflow, which in our current implementation take a few minutes (DIR and scatter correction) or an hour (the MC dose calculations), might be brought down to a few minutes altogether using fast GPU implementations, e.g. [18,19], on a single integrated platform. In such an environment, the dose calculation might also run in the background while the contours are revised to further speed up the process. Since the interactive step to find the optimal shift is very fast (less than a minute) we estimate a time benefit compared to a full adaptive re-planning, as has been discussed in [11]. Under this assumption, the dose guided approach might also have its place in integrated systems alongside with online re-planning capability. Such a workflow might look as follows:

- Acquire CBCT images
- Perform scatter correction
- Stitch pCT regions outside of scCBCT FOV to scCBCT

**Table 3**

Mean values of DVH statistics over all available fractions. The bold values indicate improvements compared to the rigid registration based shifts. For patient 1, the dose to the parotid glands was improved. For patient 2, the target coverage of the high dose PTV was increased.

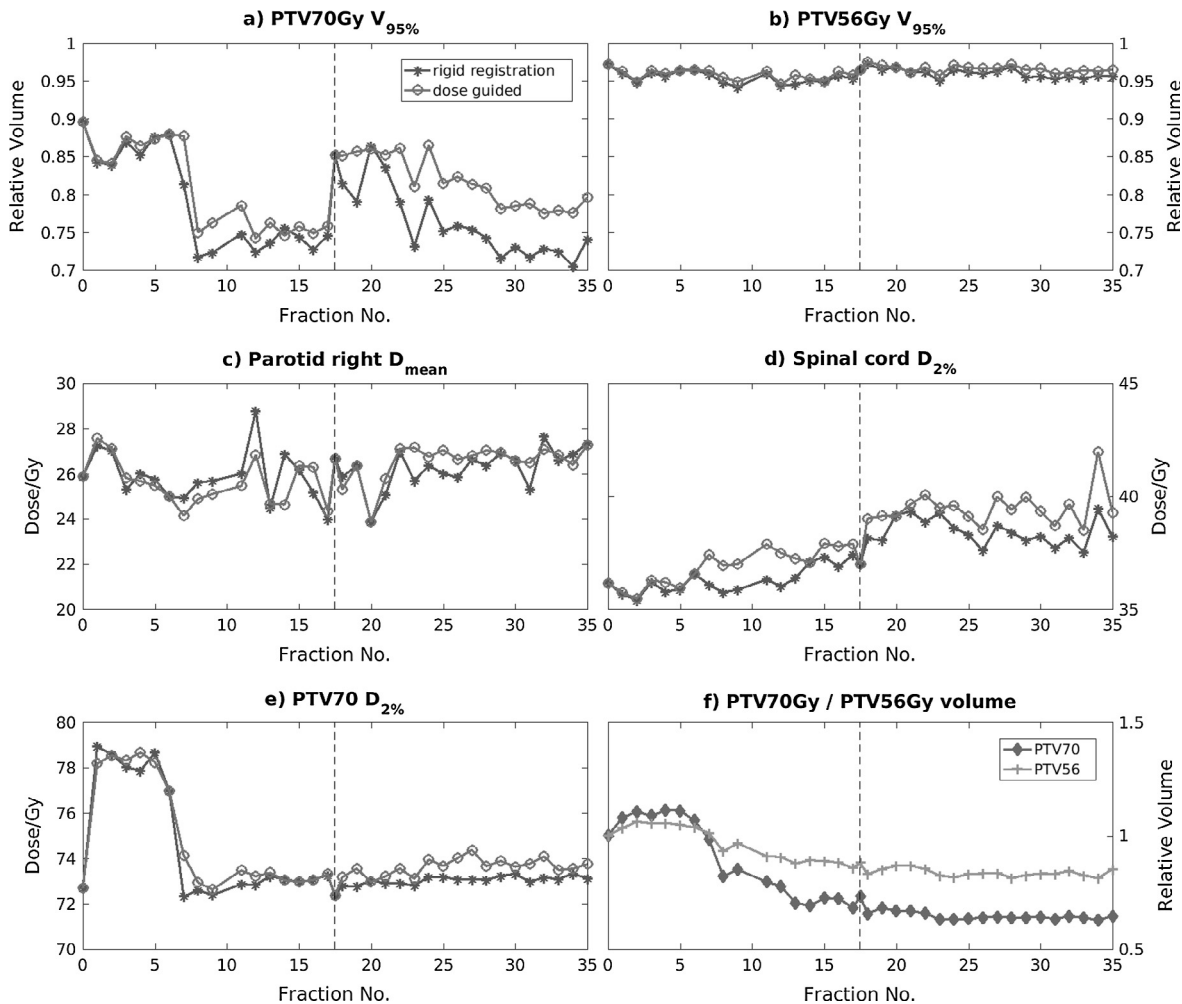
Parameter	High dose PTV V <sub>95%</sub>	Intermediate dose PTV V <sub>95%</sub>	Low dose PTV V <sub>95%</sub>	Parotid R D <sub>mean</sub> [Gy]	Parotid L D <sub>mean</sub> [Gy]	Spinal Cord D <sub>2%</sub> [Gy]	Brain stem D <sub>2%</sub> [Gy]	Opt. nerve l D <sub>2%</sub> [Gy]	Opt. nerve r D <sub>2%</sub> [Gy]	Opt. chiasm D <sub>2%</sub> [Gy]
<i>Patient 1</i>										
-planning CT	90.7%	90.7%	91.0%	29.8	29.8	36.7				
-rigid reg.	90.4%	89.5%	91.6%	<b>31.0</b>	<b>26.4</b>	38.6	n. a.	n. a.	n.a.	n.a.
-dose guided	89.9%	89.1%	91.4%	<b>30.2</b>	<b>25.3</b>	38.1				
<i>Patient 2</i>										
-planning CT	90%		97%	25.9	57.5	36.2	39.7	48.1	29.9	39.8
-rigid reg.	<b>77%</b>	n.a.	96%	26.1	57.7	37.4	40.1	49.2	29.4	39.2
-dose guided	<b>81%</b>		96%	26.1	58.1	38.3	41.2	49.2	30.0	39.8



**Fig. 2.** DVHs for one fraction of patient 1. The DVHs are scaled to the total plan dose. Compared to the rigid image registration based shifts, the dose guided approach showed comparable DVHs for PTV66Gy, PTV60Gy, PTV54Gy and the spinal cord. The dose to the left parotid gland was decreased; the dose to the right parotid gland showed a small increase.

- Revise contours
- Perform dose guided alignment
- In case a satisfactory isocenter correction is found, apply treatment
- If not, trigger adaptive re-planning (including revision/approval of the adapted plan and QA)

Dose guided positioning could have the potential to reduce the mean time per fraction compared to a daily re-planning, while still preserving DVH objectives. In clinical practice, one might also think about monitoring DVH parameters during the fractions only using dose guided positioning, and making the decision for adaptive re-planning before the next fraction as soon as parameters of interest come close to pre-defined thresholds. Methods to identify patients who benefit the most from adaptive radiotherapy might also help guiding decision making [20,21]. As already discussed in [11], a potential advantage of the proposed workflow over a full adaptive re-planning is that no new plan is created, which would always require the approval of a senior physician and a physicist. The approach has therefore the potential to reduce the required manpower during patient treatment. Furthermore, there is no need for quality assurance of a new plan. To date it is unclear how many full re-plannings for H&N patients are actually beneficial, but there



**Fig. 3.** Evolution over time of selected DVH parameters (a-e) and PTV volumes (f) for patient 2. The calculated values are the values for the corresponding fraction and on scCBCT, scaled to the total plan dose. The dashed vertical line indicates the re-planning. The planning CT is included as day 0, the re-planning CT is included between day 17 and day 18. PTV volumes are normalized to their volumes on the planning CT.



are data suggesting that for many patients only 3 adaptive replannings might be sufficient [22].

In principle, dose guided positioning might also be used in integrated systems with magnetic resonance imaging (MRI) capability, given that an accurate dose calculation on MRI is available, e.g. on synthetic CTs [23,24].

An important issue regarding a potential clinical implementation of the approach is the overall uncertainty of the DVH parameters, which is generally difficult to address. While the scatter correction approach can overcome uncertainties in the pCT to CBCT DIR and has been successfully validated in [9], and the influence of the pCT stitching to the uncertainty is considered negligible, as discussed above, the uncertainty of the updated contours depends strongly on the individual physician re-contouring the scCBCTs. Inter- and intra-observer variability in the contouring is difficult to quantify and hence to include into the evaluation of feasibility studies. Since all scCBCTs for a patient in this study have been contoured by the same physician, it is only influenced by the – potentially smaller – intra-observer variability. However, since the potential improvements of 3–5% have to be set in proportion to the overall uncertainty, this issue currently remains a major hurdle for a clinical implementation of the approach, as it is the case for any procedure with the need for updated contours.

An aspect which was not evaluated in the present feasibility study is the potential for margin reductions. In this study the coverage of the PTV was evaluated. In case a CTV can be determined with sufficient confidence based on the image of the day, it might be more plausible to use the CTV coverage in the multicriterial optimization, or a reduced PTV with smaller margins than the ones used in rigid registration based workflows.

## Conclusion

Dose guided patient positioning using scatter corrected CBCT images is feasible and offers increased control over target coverage and OAR dose compared to the clinical anatomy-based registration approach. In an integrated workflow alongside with adaptive replanning, the approach could help reducing the number of full re-plannings and therefore reduce treatment time and workload compared to a daily re-planning scenario.

## Conflict of interest

The authors declare no conflict of interest.

## Acknowledgements

DFG cluster of excellence: Munich-Centre for Advanced Photonics (MAP) EXC 158.

BMBF-SPARTA (grant number 01IB31001).

Markus Alber for providing the scriptable MC dose engine MCVerify.

Larissa Ermoschkin for contouring organs-at-risk.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.radonc.2017.09.020>.

## References

- [1] Jaffray DA, Siewerdsen JH, Wong JW, Martinez AA. Flat-panel cone-beam computed tomography for image-guided radiation therapy. *Int J Radiat Oncol Biol Phys* 2002;53:1337–49. [https://doi.org/10.1016/S0360-3016\(02\)02884-5](https://doi.org/10.1016/S0360-3016(02)02884-5).

- [2] Oldham M, Létourneau D, Watt L, Hugo G, Yan D, Lockman D, et al. Cone-beam-CT guided radiation therapy: A model for on-line application. *Radiother Oncol* 2005;75:1–8. <https://doi.org/10.1016/j.radonc.2005.03.026>.
- [3] Shi W, Li JG, Zlotecki RA, Yeung A, Newlin H, Palta J, et al. Evaluation of kV cone-beam CT performance for prostate IGRT. *Am J Clin Oncol* 2011;34:16–21. <https://doi.org/10.1097/COC.0b013e3181d26b1a>.
- [4] Barker JL, Garden AS, Ang KK, O'Daniel JC, Wang H, Court LE, et al. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. *Int J Radiat Oncol Biol Phys* 2004;59:960–70. <https://doi.org/10.1016/j.ijrobp.2003.12.024>.
- [5] Müller BS, Duma MN, Kampfer S, Nill S, Oelfke U, Geinzt H, et al. Impact of interfractional changes in head and neck cancer patients on the delivered dose in intensity modulated radiotherapy with protons and photons. *Phys Medica* 2015;31:266–72. <https://doi.org/10.1016/j.ejmp.2015.02.007>.
- [6] Cheung J, Aubry JF, Yom SS, Gottschalk AR, Celi JC, Pouliot J. Dose recalculation and the Dose-Guided Radiation Therapy (DGRT) process using megavoltage cone-beam CT. *Int J Radiat Oncol Biol Phys* 2009;74:583–92. <https://doi.org/10.1016/j.ijrobp.2008.12.034>.
- [7] Niu T, Sun M, Star-Lack J, Gao H, Fan Q, Zhu L. Shading correction for on-board cone-beam CT in radiation therapy using planning MDCT images. *Med Phys* 2010;37:5395–406. <https://doi.org/10.1118/1.3483260>.
- [8] Park Y, Sharp GC, Phillips J, Winey BA. Proton dose calculation on scatter-corrected CBCT image: feasibility study for adaptive proton therapy. *Med Phys* 2015;42:4449–59. <https://doi.org/10.1118/1.4923179>.
- [9] Kurz C, Kamp F, Park Y-K, Zöllner C, Rit S, Hansen D, et al. Investigating deformable image registration and scatter correction for CBCT-based dose calculation in adaptive IMPT. *Med Phys* 2016;43:5635–46. <https://doi.org/10.1118/1.4962933>.
- [10] 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;40:51714. <https://doi.org/10.1118/1.4801910>.
- [11] Haehnle J, Süß P, Landry G, Teichert K, Hille L, Hofmaier J, et al. A novel method for interactive multi-objective dose-guided patient positioning. *Phys Med Biol* 2017;62. <https://doi.org/10.1088/1361-6560/62/1/165>.
- [12] Haber E, Modersitzki J. Intensity gradient based registration and fusion of multi-modal images. *Methods Inf Med* 2007;46:292–9. <https://doi.org/10.1160/ME9046>.
- [13] Fischer B, Modersitzki J. Curvature based image registration. *J Math Imaging Vis* 2003;18:81–5. <https://doi.org/10.1023/A:1021897212261>.
- [14] Thieke C, Küfer KH, Monz M, Scherrer A, Alonso F, Oelfke U, et al. A new concept for interactive radiotherapy planning with multicriteria optimization: first clinical evaluation. *Radiother Oncol* 2007;85:292–8. <https://doi.org/10.1016/j.radonc.2007.06.020>.
- [15] Craft DL, Hong TS, Shih HA, Bortfeld TR. Improved planning time and plan quality through multicriteria optimization for intensity-modulated radiotherapy. *Int J Radiat Oncol Biol Phys* 2012;82:83–90. <https://doi.org/10.1016/j.ijrobp.2010.12.007>.
- [16] Kierkels RGJ, Visser R, Bijl HP, Langendijk JA, van't Veld AA, Steenbakkers RJHM, et al. Multicriteria optimization enables less experienced planners to efficiently produce high quality treatment plans in head and neck cancer radiotherapy. *Radiat Oncol* 2015;10:87. <https://doi.org/10.1186/s13014-015-0385-9>.
- [17] Thor M, Owosho AA, Clark HD, Oh JH, Riaz N, Hovan A, et al. Internal and external generalizability of temporal dose response relationships for xerostomia following IMRT for head and neck cancer. *Radiother Oncol* 2016;122:200–6. <https://doi.org/10.1016/j.radonc.2016.11.005>.
- [18] Gu X, Pan H, Liang Y, Castillo R, Yang D, Choi D, et al. Implementation and evaluation of various demons deformable image registration algorithms on a GPU. *Phys Med Biol* 2010;55:207–19. <https://doi.org/10.1088/0031-9155/55/1/012>.
- [19] Gu X, Choi D, Men C, Pan H, Majumdar A, Jiang SB. GPU-based ultra-fast dose calculation using a finite size pencil beam model. *Phys Med Biol* 2009;54:6287. <https://doi.org/10.1088/0031-9155/54/20/017>.
- [20] Brown E, Owen R, Harden F, Mengersen K, Oestreich K, Houghton W, et al. Predicting the need for adaptive radiotherapy in head and neck cancer. *Radiother Oncol* 2015;116:57–63. <https://doi.org/10.1016/j.radonc.2015.06.025>.
- [21] Brouwer CL, Steenbakkers RJHM, van der Schaaf A, Sopacua CTC, van Dijk LV, Kierkels RGJ, et al. Selection of head and neck cancer patients for adaptive radiotherapy to decrease xerostomia. *Radiother Oncol* 2016;120:36–40. <https://doi.org/10.1016/j.radonc.2016.05.025>.
- [22] Zhang P, Simon A, Rigaud B, Castelli J, Ospina Arango JD, Nassef M, et al. Optimal adaptive IMRT strategy to spare the parotid glands in oropharyngeal cancer. *Radiother Oncol* 2016;120:41–7. <https://doi.org/10.1016/j.radonc.2016.05.028>.
- [23] Guerreiro F, Burgos N, Dunlop A, Wong K, Petkar I, Nutting C, et al. Evaluation of a multi-atlas CT synthesis approach for MRI-only radiotherapy treatment planning. *Phys Medica* 2017;35:7–17. <https://doi.org/10.1016/j.ejmp.2017.02.017>.
- [24] Farjam R, Tyagi N, Veeraraghavan H, Apte A, Zakian K, Hunt MA, et al. Multi-atlas approach with local registration goodness weighting for MRI-based electron density mapping of head and neck anatomy. *Med Phys* 2017;44:3706–17. <https://doi.org/10.1002/mp.12303>.

# Measurement-based range evaluation for quality assurance of CBCT-based dose calculations in adaptive proton therapy

Sebastian Nepp<sup>1a)\*</sup> and Christopher Kurz

*Department of Radiation Oncology, University Hospital, LMU Munich, 81377 Munich, Germany*

*Department of Medical Physics, Faculty of Physics, Ludwig-Maximilians-Universität München (LMU Munich), 85748 Garching bei München, Germany*

Daniel Köpl<sup>2\*\*</sup> and Indra Yohannes<sup>3\*\*\*</sup>

*Rinecker Proton Therapy Center, 81371 Munich, Germany*

Moritz Schneider and David Bondesson

*Department of Radiology, University Hospital, LMU Munich, 81377 Munich, Germany*

*Comprehensive Pneumology Center Munich (CPC-M), German Center for Lung Research (DZL), 81377 Munich, Germany*

Moritz Rabe

*Department of Radiation Oncology, University Hospital, LMU Munich, 81377 Munich, Germany*

*Department of Medical Physics, Faculty of Physics, Ludwig-Maximilians-Universität München (LMU Munich), 85748 Garching bei München, Germany*

Claus Belka

*Department of Radiation Oncology, University Hospital, LMU Munich, 81377 Munich, Germany*

*German Cancer Consortium (DKTK), Partner site Munich, 81377 Munich, Germany*

Olaf Dietrich

*Department of Radiology, University Hospital, LMU Munich, 81377 Munich, Germany*

Guillaume Landry

*Department of Radiation Oncology, University Hospital, LMU Munich, 81377 Munich, Germany*

*Department of Medical Physics, Faculty of Physics, Ludwig-Maximilians-Universität München (LMU Munich), 85748 Garching bei München, Germany*

Katia Parodi

*Department of Medical Physics, Faculty of Physics, Ludwig-Maximilians-Universität München (LMU Munich), 85748 Garching bei München, Germany*

Florian Kamp

*Department of Radiation Oncology, University Hospital, LMU Munich, 81377 Munich, Germany*

(Received 17 July 2020; revised 8 April 2021; accepted for publication 10 May 2021; published 29 June 2021)

**Purpose:** The implementation of volumetric in-room imaging for online adaptive radiotherapy makes extensive testing of this image data for treatment planning necessary. Especially for proton beams the higher sensitivity to stopping power properties of the tissue results in more stringent requirements. Current approaches mainly focus on recalculation of the plans on the new image data, lacking experimental verification, and ignoring the impact on the plan re-optimization process. The aim of this study was to use gel and film dosimetry coupled with a three-dimensional (3D) printed head phantom (based on the planning CT of the patient) for 3D range verification of intensity-corrected cone beam computed tomography (CBCT) image data for adaptive proton therapy.

**Methods:** Single field uniform dose pencil beam scanning proton plans were optimized for three different patients on the patients' planning CT (planCT) and the patients' intensity-corrected CBCT (scCBCT) for the same target volume using the same optimization constraints. The CBCTs were corrected on projection level using the planCT as a prior. The dose optimized on planCT and recalculated on scCBCT was compared in terms of proton range differences (80% distal fall-off, recalculation). Moreover, the dose distribution resulting from recalculation of the scCBCT-optimized plan on the planCT and the original planCT dose distribution were compared (simulation). Finally, the two plans of each patient were irradiated on the corresponding patient-specific 3D printed head phantom using gel dosimetry inserts for one patient and film dosimetry for all three patients. Range differences were extracted from the measured dose distributions. The measured and the simulated range differences were corrected for range differences originating from the initial plans and evaluated.

**Results:** The simulation approach showed high agreement with the standard recalculation approach. The median values of the range differences of these two methods agreed within 0.1 mm and the

interquartile ranges (IQRs) within 0.3 mm for all three patients. The range differences of the film measurement were accurately matching with the simulation approach in the film plane. The median values of these range differences deviated less than 0.1 mm and the IQRs less than 0.4 mm. For the full 3D evaluation of the gel range differences, the median value and IQR matched those of the simulation approach within 0.7 and 0.5 mm, respectively. scCBCT- and planCT-based dose distributions were found to have a range agreement better than 3 mm (median and IQR) for all considered scenarios (recalculation, simulation, and measurement).

**Conclusions:** The results of this initial study indicate that an online adaptive proton workflow based on scatter-corrected CBCT image data for head irradiations is feasible. The novel presented measurement- and simulation-based method was shown to be equivalent to the standard literature recalculation approach. Additionally, it has the capability to catch effects of image differences on the treatment plan optimization. This makes the measurement-based approach particularly interesting for quality assurance of CBCT-based online adaptive proton therapy. The observed uncertainties could be kept within those of the registration and positioning. The proposed validation could also be applied for other alternative in-room images, e.g. for MR-based pseudoCTs. © 2021 The Authors. *Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine. [<https://doi.org/10.1002/mp.14995>]

Key words: adaptive proton therapy, gel dosimetry, range evaluation, scatter corrected CBCT

## 1. INTRODUCTION

Although the concept of adaptive radiotherapy was already described for photons in 2008 by Di Yan et al.,<sup>1</sup> real online adaptive strategies started to be increasingly used in radiation therapy clinics only in the last years. This was mainly triggered by the recently commercially available integration of magnetic resonance imaging in linear accelerators (MRI linacs).<sup>2,3</sup> However, the potential to increase dose delivery accuracy, reduce treatment margins, increase treatment efficacy, and enable new treatment concepts<sup>4</sup> makes online adaptation an attractive option for conventional linac radiotherapy as well as for proton therapy facilities. This entails making use of on-board cone beam computed tomography (CBCT) devices typically used for patient positioning.

Online replanning relies on volumetric image data acquired in the treatment room. MRI, CBCT and x-ray computed tomography (CT)-on-rails have the potential to fulfill the necessary requirements. Except for the latter, these imaging modalities do not have the intrinsic quality or information to allow for sufficiently accurate dose calculation,<sup>5</sup> especially in the case of proton therapy. CBCT is widely used in photon and proton radiotherapy departments, and the implementation of a CBCT-based online adaptive workflow would largely increase the availability of online adaptive radiotherapy. The higher sensitivity of protons to inaccuracies in the planning images compared to photons suggests that a measurement-based analysis of the image data used for dose optimization would increase confidence in the online adaptive process.

Compared to CT imaging, additional artifacts appear for CBCT imaging. These are based on the cone beam divergence, increased scatter contribution and the higher noise level.<sup>6</sup> To allow for accurate proton dose calculation, corrections must be applied to the CBCT images and extensive testing including experimental verification is necessary before implementing CBCT for online adaptive proton therapy.

Several approaches have shown to allow proton dose calculations on CBCT.<sup>7–19</sup> Dose-based quality assessment of CBCT images is usually performed by comparing recalculations of the initial plan on pairs of anatomically matched CT and corrected CBCTs. For such an analysis the time delay between the CT and the CBCT acquisition should be small compared to the time scale of expected anatomical changes of the corresponding treatment site. Therefore, in regions with small anatomical changes like the brain, the planning CT may be used, while in other regions a replanning CT or a CT which is deformably registered to the CBCT, is needed to represent the anatomy of the day. The dose distribution originating from a forward calculation of the plan on the CBCT is then compared to a dose calculation of the same plan on the anatomically matched CT using dose difference or gamma pass rates. Additionally, range differences in beam-eye-view are typically analyzed.<sup>8–10,13,14,16</sup>

Methods to verify experimentally the results obtained by recalculation are currently missing. The published studies only evaluate the range differences based on an in-silico analysis, while this study makes the influence of the CBCT image quality on the dose calculation measurable. For this, the actual usage of the image data in an adaptive treatment workflow must be reproduced. Irradiation of two different plans on a patient-specific phantom, one optimized on CT and the other on CBCT images, allows for a measurement-based range comparison. For this comparison, measurements of two-dimensional (2D) or even three-dimensional (3D) dose distributions with an adequate spatial resolution are necessary. While film dosimetry allows for accurate 2D dose and range determination,<sup>20</sup> Hillbrand et al. showed that gel dosimetry (using MR for readout) in combination with a 3D printed head phantom can be used to determine proton range in 3D accurately.<sup>21</sup>

In this contribution, we introduce a measurement-based quality assurance tool to evaluate the use of corrected CBCT



image data for online adaptive proton treatment planning. The tool is based on a range comparison of measured dose distributions for single field uniform dose (SFUD) pencil beam scanning proton plans. Three water-filled 3D-printed, patient-derived head phantoms with inserts for polymer gel and film dosimetry were used for measurements of 3D and 2D dose distributions respectively. To measure the impact of CBCT image quality on dose optimization and calculation without introducing additional anatomical variations between the CBCT and the corresponding high-quality CT, we used data from brain cancer cases in this study. Since no major deformations are expected in the brain, the study design allowed to focus on the pixel-wise mismatch of CT values and reduce anatomical variations to a small uncertainty originating from a rigid registration. The measured and calculated dose distributions are compared based on an evaluation of range difference maps (3D) and range difference curves (2D).

## 2. MATERIALS AND METHODS

### 2.A. Data preparation

In preparation for the study, three patients were selected who underwent photon radiation therapy and had a planning CT scan with a voxel size of  $1.074 \times 1.074 \times 1 \text{ mm}^3$ . Brain cancer cases were selected in this project to avoid introducing anatomical variations between CT and CBCT images. The CT scans were acquired with a Toshiba Aquilion Large Bore CT scanner (Canon Medical Systems, Japan) in a standard radiation therapy setup, using a flat tabletop and an immobilization mask. These initial planning CT scans are denoted by planCT in the following. Additionally, CBCT images were acquired as part of clinical routine with the standard clinical head protocol (100 kV, 10 mA, 10 ms,  $200^\circ$ , 197–209 frames) using the XVI 5.0.4 on-board x-ray imager of an Elekta Synergy linear accelerator (Elekta Ltd, Sweden). Planning CT scans and CBCT acquisitions were acquired at the LMU University Hospital Munich, while treatment planning and irradiations were performed at the Rinecker Proton Therapy Center (RPTC) in Munich.

The CBCT images were corrected following the approach used in Kurz et al.,<sup>10</sup> which is based on the work of Park et al.<sup>15</sup> and Niu et al.<sup>22</sup> and was shown to correct for scatter and beam hardening.<sup>23</sup> The calibration factor, which is used to scale the primary projections, was set to 25.6. The corrected CBCT images are denoted with scCBCT in the following.

To account for differences in the Hounsfield unit (HU) calibration of the CT scanner used for patient data acquisition and the Philips Brilliance 16 P CT scanner (Philips, Netherlands) used for treatment planning at RPTC, a mapping was performed. For this, the Gammex CT Electron Density Phantom (Sun Nuclear Corporation, United States) was used to correlate the HU values of both scanners. A piecewise linear conversion function was defined among the 13 data point resulting from the different inserts with defined electron density in the phantom.

### 2.B. Phantom setup

Based on the bony structure and outer contour extracted from the CT scans of the three selected patients, three different head phantoms were 3D printed (individualized Prime phantoms by RTSafe P.C., Greece). One printing material was used as skin and bone surrogate. The printed head was filled with water to mimic soft tissue. The head design was changed compared to the one presented in Markis et al.<sup>24</sup> allowing to insert different detectors into a circular opening with a diameter of 80 mm at the caudal end of the heads. A second smaller hole can be used to fill the phantom with water (see Fig. 1). Due to limitations in the agreement between the CT scan and the resulting 3D printed head phantom, related to limitations of printing fine structures, one single printing material and reproducibility, a study design using one reference head per patient was preferred over printing multiple heads for each patient. For this project, gel cylinders (inner dimensions: 74 mm diameter, 140 mm height) filled with an N-vinylpyrrolidone-based polymer dosimetry gel<sup>25</sup> and a film holder equipped with pieces of Gafchromic EBT3 film (Ashland Inc., United States) were used as inserts. The film holder was made of RW3 solid water. An exemplary axial slice of the head phantom for patient 1 is presented with the film holder in Fig. 2(c) and with the gel cylinder in Fig. 2(d). Four nails connecting the two holding plates fix the film. Then the two joined plates are pushed into the film holder slot insert, which is fixed to the head with screws. This design allows for quick changes of the film pieces without having to refill water. The nails in the film holder enable a landmark-based registration between the films and a CT scan of the head phantom with the film holder. Additionally, four

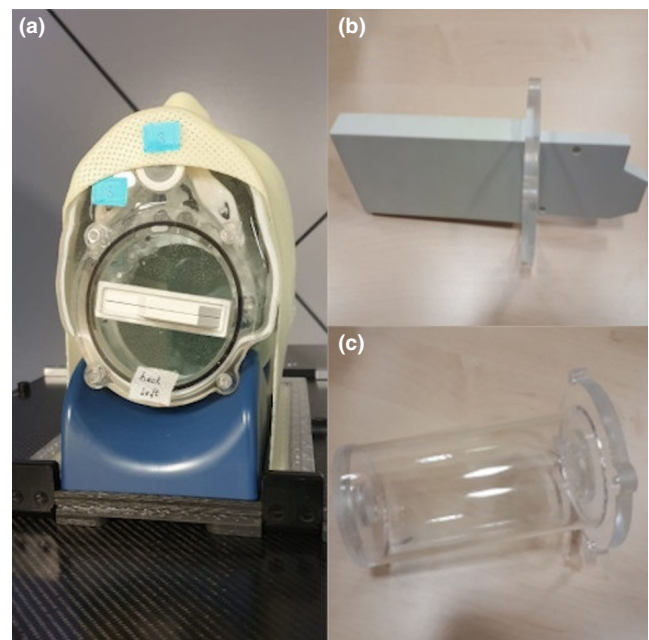


FIG. 1. Photo of a head phantom fixed with pillow and thermoplastic mask (a). The film holder is mounted in the detector opening. Additionally, the film holder (b) and an empty gel cylinder (c) are presented. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

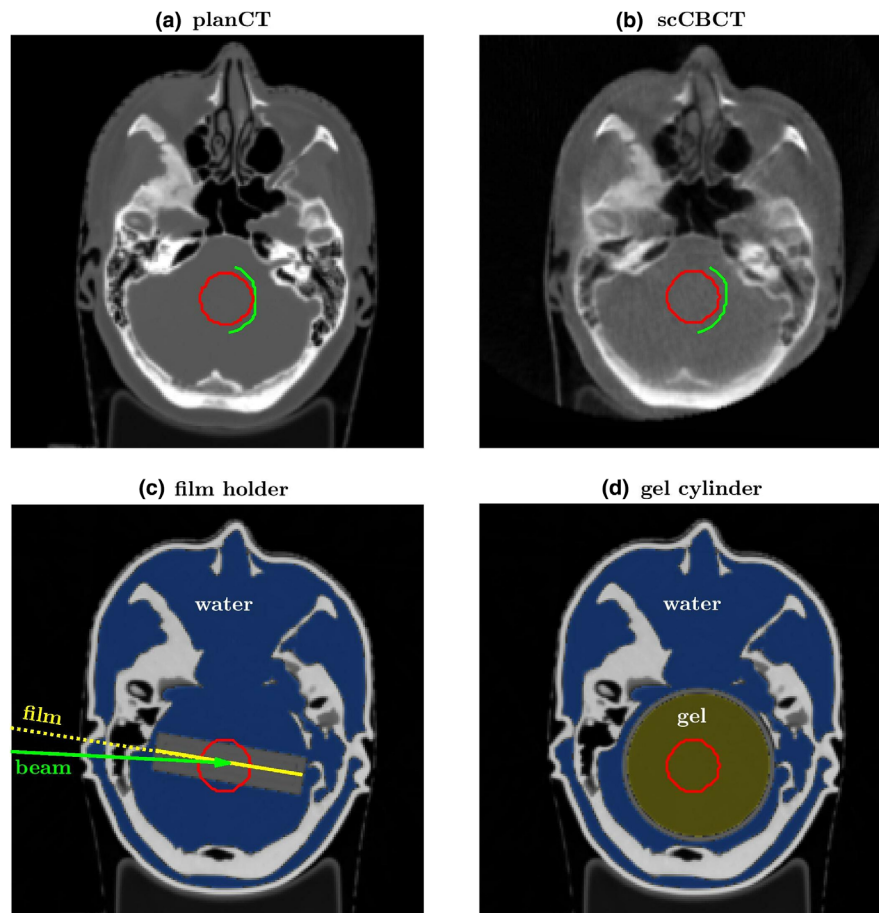


FIG. 2. Exemplary axial slice of patient 1 of the planCT (a), the scCBCT (b), the 3D printed head with film holder (c) and with gel cylinder (d). The target volume is drawn in all subfigures with a red line. In subfigures a and b the distal 80% fall-off is indicated by a green line for  $NP_{\text{planCT}}$  and  $RP_{\text{scCBCT}}$  respectively. In subfigure c, the film plane and beam direction are highlighted showing the angular deviation of  $\sim 6^\circ$  in between. The water filling is shown in blue and the gel in the PMMA cylinder is marked in olive-green (d). [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

MR PinPoint<sup>®</sup> markers (Par Scientific A/S, Denmark) were glued to each of the heads to simplify registrations with the MR scans used for reading out the dosimetry gel.

The relative stopping power ratios (RSPs) for the film holder and the gel were determined in a previous experiment using ranges acquired with PTW PEAKFINDER (PTW, Germany). For the film holder an RSP of 1.035 was calculated using a measurement with and without the film holder in the beam. For the gel, similar measurements were performed with a box made of polymethyl methacrylate filled with water or dosimetry gel. The resulting RSP was  $1.025 \pm 0.001$  (mean and one standard deviation testing three gel batches) respectively. According to van Abbema et al.<sup>26</sup> the uncertainty in the determination of the RSP based on length and range measurement is below 0.4%. Since no sample length measurement is necessary for the RSP measurement of the gel, the uncertainty is deemed even lower and 0.4% was assumed as an upper limit. Since no dose calculations were performed on phantom CT scans, the RSP determination of the film holder and the gel was only needed to ensure that the different setup did not lead to a range shift in the comparison of gel and film measurements. The small observed deviation of the RSP values of gel and film was considered to be

negligible for the range comparison between the two measurement methods. Additionally, ranges were not compared directly between gel and film measurements, but only as relative range differences (in the order of few mm).

The phantom heads were placed on positioning pillows and fixed with thermoplastic masks. CT scans of the phantoms with fixation and without water filling were acquired with the film holder and an empty gel cylinder for target definition and as a reference for registration of the patient CT and CBCT scans.

## 2.C. Adaptive treatment planning

The initial treatment plan was created on the planCT (patient CT scans) for a cylindrical target volume with a diameter of 30 and 20 mm minimum distance to the edges of the gel cylinder (see Fig. 2) using the clinical treatment planning software (TPS) XiO 4.80 (Elekta Ltd, Sweden). One SFUD pencil beam scanning proton plan from one gantry angle was created per patient. The isocenter was set to the center of the target volume and the plan was optimized for one fraction of 8 Gy, which is recommended for the used dosimetry gel. The dose grid size was set to 2 mm in all dimensions. For the



optimization, the peak width multiplier, which defines the longitudinal spot spacing,<sup>27</sup> was set to 1.2 and the lateral spot spacing to 5 mm. Following the idea of Zhao et al.<sup>20</sup> the gantry angle was chosen to have approximately 6° difference to the angle of the film holder to avoid that the beam passes the film holder through the gap between the two plates [see Fig. 2(c)]. Maximum iterations for a single field were set to 300.

For adaptive replanning, the scCBCT images were manually rigidly registered to the planCT images using the contouring software CMS Focal 4.8 (Computerized Medical Systems, United States), which was also used for contouring on the planCT. The target contour was copied from planCT to scCBCT after registration. Treatment planning was performed using the same settings and optimization constraints as for the initial treatment plan. Due to the HU calibration (see Section 2.A) of the image sets, the standard TPS calibration curve could be used.

For the evaluation, additional recalculations of the radiation therapy plans optimized on planCT (NP—nominal plan) and on scCBCT (RP—reoptimized plan) with a 1 mm dose grid were performed on scCBCT and planCT with the TPS XiO. This means that RP and NP were optimized on the same target using similar anatomical information, but with different image quality (CT vs CBCT).

## 2.D. Plan delivery

In total, six plans were delivered on each of the three heads. Per head, both NP and RP were irradiated on two different film pieces and for patient 1 on one gel cylinder. A gel dosimetry measurement was only performed for the plans of the first patient for financial and logistic reasons. Prior to irradiation, the film holder was mounted in the phantom head, which was in turn filled with water at treatment room temperature and positioned on the treatment table with pillow and mask. Positioning was verified with an orthogonal x-ray imager mounted to the gantry and corrections were performed according to manual registrations. After the irradiation of both film pieces, the first gel cylinder was inserted in the head phantom. Due to the design of the phantom, water had to be refilled and x-ray positioning was necessary for every gel cylinder change.

## 2.E. Dose readout

The 3D dose distribution in the gel was extracted with an MR scanner. The readout of the gel cylinders (mounted in the water-filled heads) was performed at a Magnetom Aera 1.5-Tesla whole-body MR scanner with a 20-channel head-and-neck coil (Siemens Healthineers, Germany) about 48 h after irradiation. Four independent asymmetric spin echoes were acquired after each excitation at echo times of TE = 36, 436, 835, 1230 ms with a multicontrast half-Fourier-acquired single-shot turbo-spin-echo (HASTE) sequence. For more details and other sequence parameters see Hillbrand et al.<sup>21</sup>

The acquired multiecho image data were converted to quantitative  $R_2$  values ( $R_2 = 1/T_2$ , transverse relaxation rate) in a postprocessing step.<sup>21,28,29</sup> The  $R_2$  values are proportional to the delivered dose and therefore allow for range determination. MR scanning took about 40 min per gel cylinder for 75 slices with a slice thickness of 3 mm. Pixel spacing was set to  $0.68 \times 0.68$  mm<sup>2</sup> within the axial slices. The water temperature in the head was measured at each cylinder replacement to avoid an increase of the gel temperature above the tolerances of the gel.

For the film readout, an Epson Expression 11000 XL Pro (Epson, Japan) flatbed scanner was used. To reduce potential influences of differences in illumination due to slightly different position of the films, a cardboard template was used to position the film on the scanner and removed before starting the scanning process. All corrections were turned off in the scanner software and the resolution was set to 150 dpi with a bit depth of 48. Since the pixel size is much smaller for the film compared to the other datasets, a 2D convolution with a uniform filter of size  $5 \times 5$  pixel was applied to reduce the noise. For evaluation, the average of the two film scans per plan was considered.

Since only relative dose values were required in the subsequent range analysis and due to limited availability of the proton beam, the film calibration was performed using a 6 MV photon beam of the Synergy linear accelerator. A total of 20 pieces of film were irradiated with 12 dose levels between 0 and 11 Gy at the depth of maximum dose. The calibration pieces were read out 44 h after irradiation and a calibration curve was fitted following the triple channel approach of Micke et al.<sup>30</sup> The final set of films was also read out about 44 h after irradiation and the previously acquired calibration curve was applied.

## 2.F. Dataset registration

For evaluation, a 2D and a 3D reference condition were defined. The 3D reference (voxel size interpolated to  $1 \times 1 \times 1$  mm<sup>3</sup>) was given by a CT scan of the head phantom with the film holder insert using the Toshiba Aquilion Large Bore CT scanner. PlanCT and scCBCT were registered manually to the 3D reference based on the bony anatomy. These registrations were applied to the corresponding planned dose distributions. The four MR markers were used to supply landmarks for the 3D rigid registration of the gel readout MR scans. The 2D reference (pixel size interpolated to  $1 \times 1$  mm<sup>2</sup>) was the film plane with the nail positions as landmarks for registration. All film pieces were aligned based on the nail positions. A registration of the nail positions in the 3D reference to the 2D reference provided the parameters needed for registration of the 3D dose distributions to the film plane. VV version 1.4<sup>31</sup> was used to generate manual registration matrices and for the definition of landmarks. The application of the registration and generation of dose volume data was performed with plati-match version 1.7.4.<sup>32</sup>

## 2.G. Data evaluation

Except for the registrations the evaluation of the data was performed using MATLAB 2019a (The MathWorks, Inc., United States). The following dose distributions were available for the evaluation of the three patients (gel measurements were performed only for patient 1):

- $NP_{\text{planCT}}$ , the dose distribution originating from the initially optimized (i.e. nominal) plan (NP) on planCT,
- $RP_{\text{scCBCT}}$ , the dose distribution resulting from the *adaptive replanning* on scCBCT (RP) using the same optimization constraints as for NP,
- $NP_{\text{scCBCT}}$ , the recalculation of NP on scCBCT,
- $RP_{\text{planCT}}$ , the recalculation of RP on planCT,
- $NP_{\text{film}}$  and  $NP_{\text{gel}}$ , the measured dose distributions acquired when irradiating NP on the phantom with film holder and gel cylinder, respectively,
- $RP_{\text{film}}$  and  $RP_{\text{gel}}$ , the measured dose distributions obtained when irradiating RP on the phantom with film holder and gel cylinder, respectively.

The measured and calculated dose distributions were analyzed based on a range difference analysis. For the range determination, the dose cubes were rotated and interpolated so that the voxel rows point in beam direction. The subsequent range analysis could then be performed on the voxel rows. The range was defined as the distance between the patient outline and the 80% distal fall-off of the voxel rows in beam's eye view direction

$$R80(\text{dose}) = x_{80\%}(\text{dose}) - x_e(\text{dose}), \quad (1)$$

where  $x_{80\%}$  is the interpolated position of the 80% distal fall-off and  $x_e$  is the position of the first pixel belonging to the patient outline structure of the evaluated pixel row. The patient outline was created both on CT and scCBCT datasets using an HU threshold of  $-600$ .

The range evaluation was performed only for profiles intersecting the target volume. To limit the influence of the 3D printing quality on the evaluation, range differences were only calculated from datasets obtained either from a measurement on the same phantom and modality (film vs film or gel vs gel) or from calculation. This means that for the comparison between measurement and calculation only range differences were considered.

Three different evaluation approaches were considered in this study:

1. The *recalculation approach*, where the dose distributions  $NP_{\text{planCT}}$  and  $NP_{\text{scCBCT}}$  are compared.
2. The *measurement approach* uses the delivery of NP and RP on a single phantom and subsequent comparison of the resulting dose distributions.
3. The *simulation approach* mimics the *measurement approach* by comparing  $NP_{\text{planCT}}$  and  $RP_{\text{planCT}}$ .

Since the latter two methods additionally include potential differences in the distal edges of the two different optimized dose distributions, which originate from the two independent optimizations (see Fig. 2), a correction

$$\Delta R_{\text{opt}} = R80(RP_{\text{scCBCT}}) - R80(NP_{\text{planCT}}) \quad (2)$$

was considered. This could also correct for effects such as the addition/removal of a pencil beam energy layer during re-optimization due to slight water equivalent thickness differences between scCBCT and planCT.

The range differences needed for the evaluations are presented in Fig. 3. The *recalculation approach* corresponds to the analysis performed in most studies<sup>8–10,13,14,16</sup> to compare the influence of different image sets on dose calculation. For this, the range differences between  $NP_{\text{planCT}}$  and  $NP_{\text{scCBCT}}$  ( $\Delta R_{\text{rec}}$ ) were determined (see Fig. 3). For better comparability with the *simulation* and *measurement approach* the sign of this range difference is changed:

$$\Delta R_{\text{rec}} = -(R80(NP_{\text{scCBCT}}) - R80(NP_{\text{planCT}})). \quad (3)$$

While a deviation in the HU directly affects the range in recalculation, this deviation is considered in the plan optimization for the *simulation* and *measurement approach*. Therefore, the effect of a HU deviation on the range is opposite for recalculation compared to simulation and measurement, thus the negative sign.

The *measurement approach* is based on the range differences between the measured ranges of the dose distribution originating from RP and NP evaluated in the beam's eye view projection of the target volume (see Fig. 3) for the film measurement

$$\widetilde{\Delta R}_{\text{film}} = R80(RP_{\text{film}}) - R80(NP_{\text{film}}) \quad (4)$$

and the gel measurement

$$\widetilde{\Delta R}_{\text{gel}} = R80(RP_{\text{gel}}) - R80(NP_{\text{gel}}). \quad (5)$$

These differences were subsequently corrected with  $\Delta R_{\text{opt}}$ , which makes the latter two range differences comparable to  $\Delta R_{\text{rec}}$ . The corrected measured range differences are obtained by

$$\Delta R_{\text{gel/film}} = \widetilde{\Delta R}_{\text{gel/film}} - \Delta R_{\text{opt}}. \quad (6)$$

While for the gel measurement the range is determined in beam direction, the range analysis for the 2D measurements is determined in the film plane, which has an angular deviation of  $6^\circ$ . The corresponding deviation of the range difference is in the order of  $1 - \cos(6^\circ) = 0.5\%$ . The resulting deviations regarding the difference in the analyzed profile should not affect the evaluation because only profiles in the film plane were compared for the 2D evaluation.

The *simulation approach*, based on plan recalculations, mimics the *measurement approach* using planCT as a virtual phantom. This means RP was recalculated on planCT and the range difference to the original dose distribution of NP was determined (compare Fig. 3)

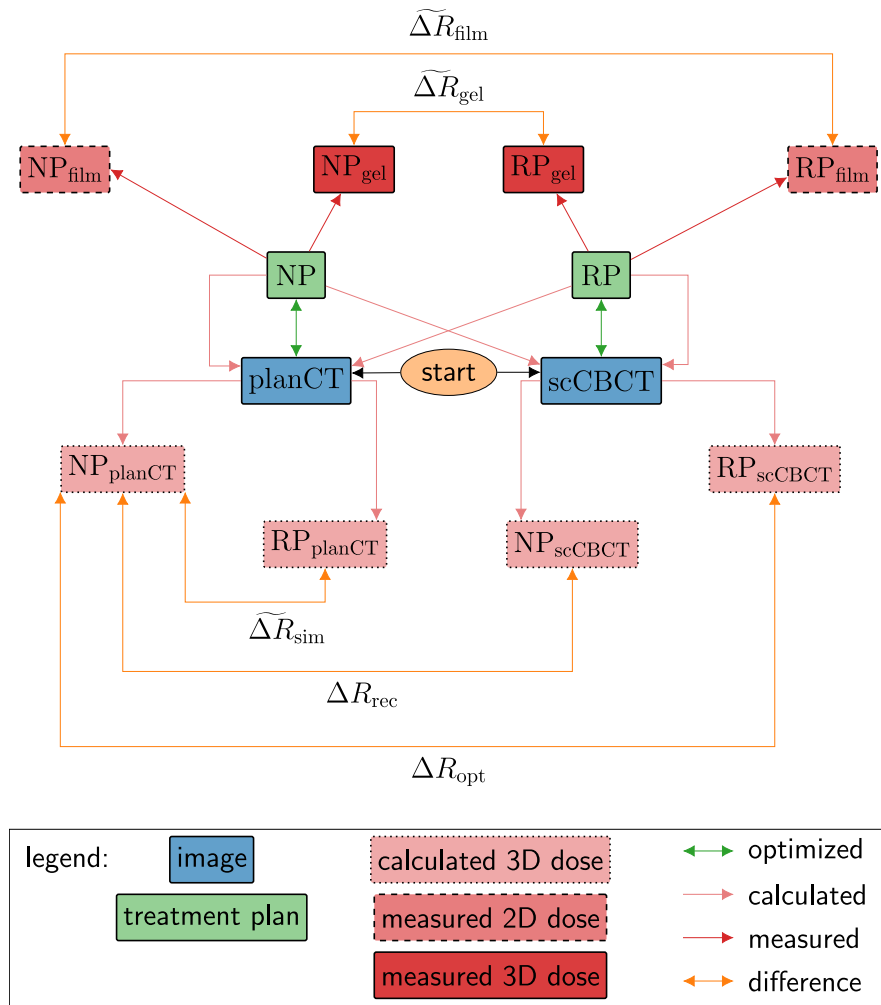


FIG. 3. Overview of the evaluation. Starting from the image data (planCT and scCBCT), two plans were optimized for each patient. The resulting plans were recalculated on the other image data and irradiated on film and dosimetry gel. The differences for ranges resulting from different plans (~ indicates uncorrected range differences) were corrected for range differences, which were already present in the initially optimized plans. [Color figure can be viewed at wileyonline library.com]

$$\widetilde{\Delta R}_{sim} = R80(RP_{planCT}) - R80(NP_{planCT}) \quad (7)$$

and corrected for range differences originating from the originally optimized dose distributions

$$\Delta R_{sim} = \widetilde{\Delta R}_{sim} - \Delta R_{opt}. \quad (8)$$

$\widetilde{\Delta R}_{sim}$  and  $\widetilde{\Delta R}_{gel/film}$  estimate the same range shift, originating from calculations on CT data and from measurements on the phantoms, respectively. These include the differences originating from differences in the optimized dose distributions when calculated on their respective image data ( $\Delta R_{opt}$ ) and differences which are based on the difference in HU values (and RSP) on the protons' path to the target.  $\Delta R_{rec}$  is only incorporating the latter since no reoptimization is needed here. However, the way the optimization is affected by HU deviations is opposite compared to recalculation. For example, a lower HU (lower RSP) value in the protons' path leads to a higher range for recalculation, while in a reoptimized plan this would be considered and lead to a plan with a lower

beam energy. Therefore,  $\Delta R_{rec}$  is the negative value of the  $\Delta R_{gel/film}$  (with  $\Delta R_{opt}$  correction). After correction of  $\widetilde{\Delta R}_{sim}$  and  $\widetilde{\Delta R}_{gel/film}$  with  $\Delta R_{opt}$  (range differences originating from the optimization), all three measures are directly comparable.

A 3 mm range difference pass rate ( $PR_{3mm}$ ) was used to analyze the calculated range difference maps. It is defined as the fraction of range differences (evaluated in beam's eye view profiles) which are smaller than 3 mm. This metric was used to analyze the optimized dose distribution on planCT and scCBCT and to compare the simulation and recalculation method.

### 3. RESULTS

#### 3.A. Treatment plan comparison

For comparison of the dose distributions from the different treatment plans ( $NP_{planCT}$ ,  $NP_{scCBCT}$  and  $RP_{scCBCT}$ ), the dose volume histogram (DVH) for the target volume of the three patients is shown in Fig. 4. The differences are negligible for

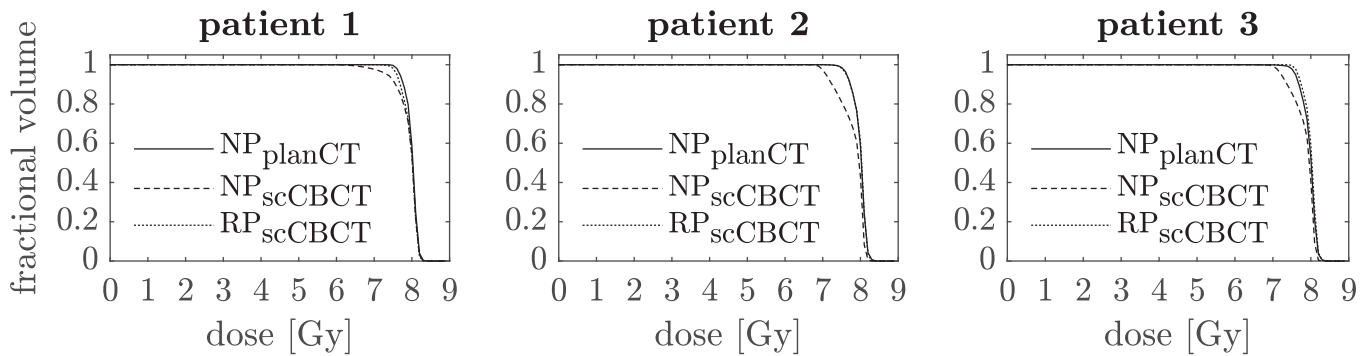


FIG. 4. Dose volume histogram comparison of the target structure for the nominal plan (NP) optimized on planCT and the reoptimized plan (RP) on scCBCT. NP was calculated both on planCT and scCBCT.

TABLE I. D98% values of the target volume for the plans optimized on planCT ( $NP_{\text{planCT}}$ ), scCBCT ( $RP_{\text{scCBCT}}$ ) and the recalculated plan on the scCBCT ( $NP_{\text{scCBCT}}$ ) are given in percent of the prescribed dose (8 Gy) for all three patients.

	Patient 1	Patient 2	Patient 3
$NP_{\text{planCT}}$	95.1%	94.1%	93.6%
$RP_{\text{scCBCT}}$	93.8%	93.9%	94.9%
$NP_{\text{scCBCT}}$	86.6%	87.0%	88.7%

the optimized plans ( $NP_{\text{planCT}}$  and  $RP_{\text{scCBCT}}$ ), while the recalculation ( $NP_{\text{scCBCT}}$ ) has a lower coverage for all three patients. This observation is supported by the dose covering 98% of the target volume (D98%), which is presented for all three patients in Table I. In Fig. 2 an exemplary axial slice of patient 1 with the planCT [Fig. 2(a)] and the scCBCT [Fig. 2(b)] including the 80% range fall-off of  $NP_{\text{planCT}}$  and  $RP_{\text{scCBCT}}$  respectively. The target structure of the respective optimized plans is shown for comparison. Differences are visible between the R80 lines for planCT and scCBCT due to differences in the optimization.

### 3.B. 2D analysis

Figure 5 presents the corrected range differences in the film plane for all three patients. For patient 1 the corrected range difference for the gel measurement was also evaluated in the film plane for comparison. Simulation and film measurement show a high agreement for all patients, with differences mainly appearing in the caudal part. The maximum deviation at a given position of  $\Delta R_{\text{film}}$  and  $\Delta R_{\text{sim}}$  was 1 mm for all three patients. The median values of  $\Delta R_{\text{film}}$  were 1.3 (1.1) mm, 1.0 (0.5) mm, and 1.9 (1.7) mm for patients 1, 2, and 3 respectively. The median values of  $\Delta R_{\text{sim}}$  were 1.4 (1.0) mm, 1.0 (0.5) mm, and 1.8 (1.3) mm for patients 1, 2, and 3 respectively. The gel measurement for patient 1 showed the same trend as observed for film measurement and simulation. The maximum deviation of 1.4 mm between gel measurement and simulation was larger than for the film measurement, and the median value of  $\Delta R_{\text{gel}}$  for patient 1 was 0.7 (1.0) mm.

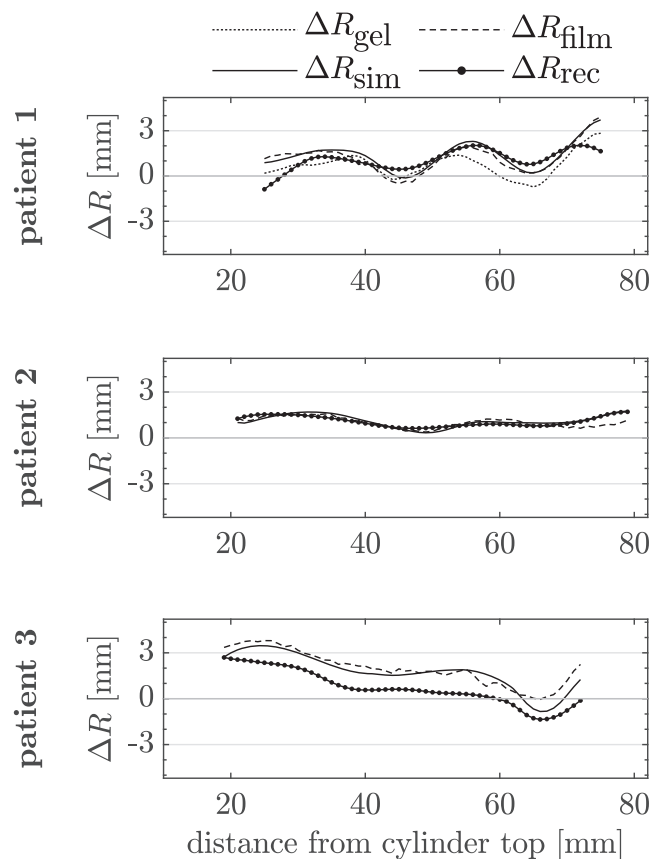


FIG. 5. The corrected range differences for all three patients evaluated in the film plane for the film measurement and simulation of the measurement on the planCT. The range difference of the recalculation is shown for comparison. For patient 1 additionally the corrected range differences for the gel measurement evaluated in the film plane are plotted.

### 3.C. 3D analysis

Table II presents  $PR_{3\text{mm}}$  and the median and IQR of  $\Delta R_{\text{sim}}$  and  $\Delta R_{\text{rec}}$ . The deviation between the pass rates is below 2% for all patients for the two methods. For patient 1, additionally the corresponding values of the gel measurement are shown. While the median of the  $\Delta R_{\text{opt}}$  correction is



almost zero for patients 1 and 2, the correction for patient 3 indicates a slight systematic variation of more than 1 mm.

Figure 6 shows 2D maps of range differences for the three evaluation approaches (*recalculation*, *simulation*, and *gel measurement*). For the *gel measurement* and the *simulation approach* the range differences were corrected for range differences in the original plans. Additionally, a projection of the head phantom in beam's eye view highlighting the evaluated area in red is presented. The three range maps show the same trends. The shape of the ear, which is in the beam line for patient 1, is visible in the corrected range maps for *measurement*, *simulation*, and *recalculation*. The visibility of the ear results from small differences in the deformation of the ear for planCT and scCBCT. The main differences appear in the caudal and posterior region of the target volume. A direct comparison between the simulation and gel measurement showed that 98.5% of the evaluated range differences of both methods for patient 1 agreed within 2 mm.

A box plot of the corrected range differences evaluated for the 3D dose distributions of patient 1 for gel measurement ( $\Delta R_{\text{gel}}$ ), simulation ( $\Delta R_{\text{sim}}$ ), and recalculation ( $\Delta R_{\text{rec}}$ ) is shown in Fig. 7. For all three methods the values are symmetrically distributed. The box plot confirms, what was observed in the range difference maps. All corrected range differences show a slight positive trend and the median of  $\Delta R$  was almost the same for both simulation and recalculation (see Table II). Only for patient 1 the IQRs differed with 1.7 mm and 1.4 mm for recalculation and simulation respectively. The evaluation of the gel ranges resulted in a median value for  $\Delta R_{\text{gel}}$  of 0.6 (1.9) mm.

#### 4. DISCUSSION

A novel measurement-based evaluation approach for proton dose calculation accuracy on scCBCT images was presented, with the concept of online adaptive proton therapy in mind. The presented quality assurance workflow allows the evaluation of the accuracy and hence feasibility of scCBCT-based online adaptive proton therapy. Range differences originating from optimization on the different image data were calculated and compared to measured range differences with gel and film dosimetry. The resulting range differences were also compared to the recalculation-based standard evaluation used by multiple groups for CBCT evaluation in proton

therapy.<sup>8–10,13,14,16</sup> Although the method was only applied for scCBCT it is in principle applicable for any in-room image data potentially used for treatment adaptation, like, e.g., MR-based pseudoCTs.

The DVH comparison of the two optimized treatment plans  $\text{NP}_{\text{planCT}}$  and  $\text{RP}_{\text{scCBCT}}$  showed only small differences for all three patients, while the D98% of the recalculated dose  $\text{NP}_{\text{scCBCT}}$  was about 7 percentage points lower compared to the optimized doses. The observed median range differences in the order of 1 to 2 mm for the recalculation lead to a drop of the D98%, which is very sensitive to range shifts especially for small volumes and highly conformal plans. This difference had to be compensated in the reoptimized plan (RP) and was one part of the resulting measured range differences between  $\text{NP}_{\text{planCT}}$  and  $\text{RP}_{\text{scCBCT}}$ .

Although the deviations were small for the DVH parameters, the ranges of the plans optimized on scCBCT and planCT differed in some profiles by more than 3 mm. This might originate from limitations in the clinical optimization settings using a spot spacing in depth of about 5 mm, which may cause a layer of pencil beams to be added or removed during plan re-optimization on scCBCT. These settings should be carefully considered in an online adaptive workflow. This, however, does not affect the main point of the study, which is to verify that the method is sensitive to the impact of the image data on the dose calculation.

The presented method can detect combined variations stemming from (a) the optimization and (b) the underlying image data. The optimization differences would not be considered in the standard *recalculation approach*, which only shows how the range of protons is affected by differences of the image data. The *simulation approach* additionally addresses the influence of image differences on the optimization itself. Since the latter is affected, among others, by the energy-layer spacing during the optimization process, the measurement and *simulation approach* were corrected for the differences originating from the original plans for improved comparability with the standard *recalculation approach*. While the  $\Delta R_{\text{opt}}$ -correction is very small (median values below 0.3 mm) for patient 1 and 2, the median value of  $\Delta R_{\text{opt}}$  is more than 1 mm for patient 3, which should not be neglected. The equivalence of the corrected *simulation* and *recalculation approach* was supported by the similarity in the 3 mm range difference pass rates (see Table II) and the 2D

TABLE II. Median, IQR, and pass rates for planCT vs scCBCT range differences of the beam's eye view profiles exhibiting differences below 3 mm ( $\text{PR}_{3\text{mm}}$ ) for the simulation (corrected for differences in the original plans), the recalculation approach and the gel measurement (only for patient 1). Additionally, the corresponding values are shown for the  $\Delta R_{\text{opt}}$  correction for all three patients.

	Patient 1			Patient 2			Patient 3		
	$\text{PR}_{3\text{mm}}$	Median	IQR	$\text{PR}_{3\text{mm}}$	Median	IQR	$\text{PR}_{3\text{mm}}$	Median	IQR
Simulation ( $\Delta R_{\text{sim}}$ )	90.1%	1.3 mm	1.4 mm	98.0%	1.1 mm	0.8 mm	92.8%	2.0 mm	0.8 mm
Recalculation ( $\Delta R_{\text{rec}}$ )	88.3%	1.3 mm	1.7 mm	97.5%	1.1 mm	0.8 mm	92.6%	2.0 mm	0.8 mm
Gel measurement ( $\Delta R_{\text{gel}}$ )	93.5%	0.6 mm	1.9 mm						
Correction ( $\Delta R_{\text{opt}}$ )	98.9%	-0.2 mm	1.3 mm	100%	0.0 mm	0.4 mm	97.6%	-1.4 mm	0.8 mm



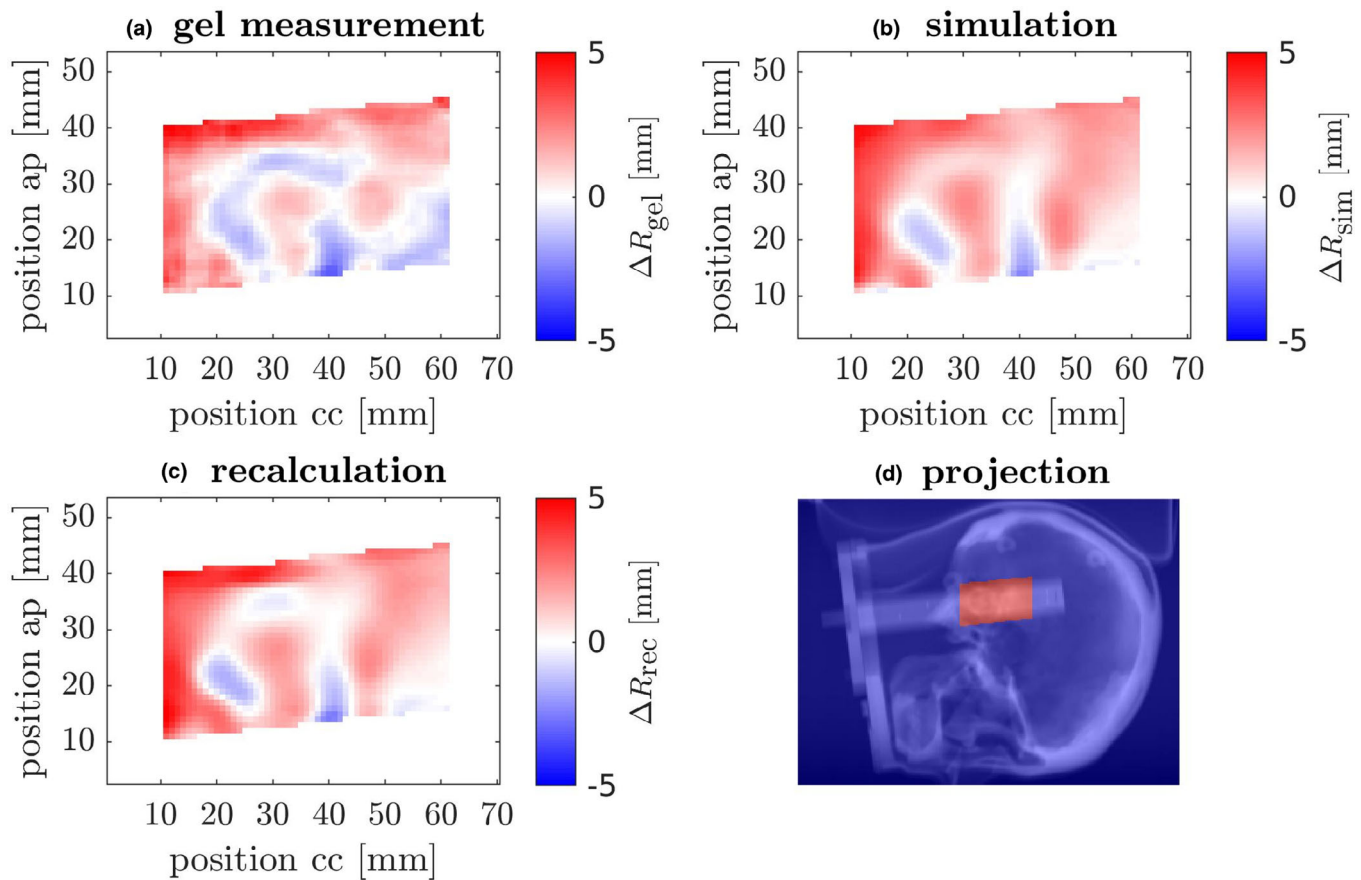


FIG. 6. Corrected range difference maps for patient 1 of the gel measurement (a), simulation (b) and the range difference map of the recalculation (c). In the projection view of the three-dimensional printed head with the film holder in beam's eye view, the evaluation area is marked in red (d). [Color figure can be viewed at wileyonlinelibrary.com]

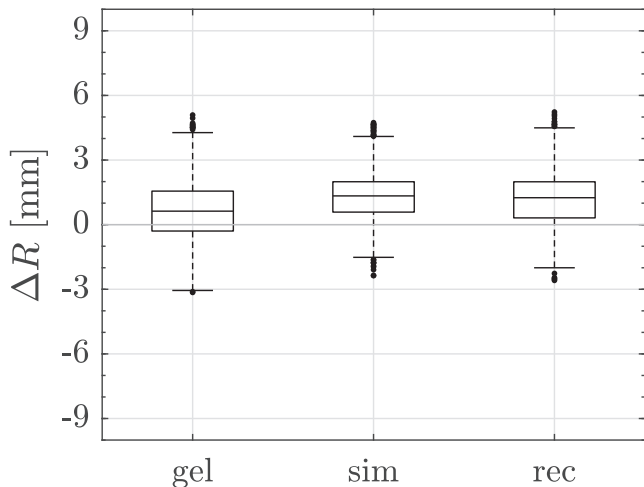


FIG. 7. Box plot of the corrected range differences of patient 1 for the gel measurement (gel), the simulation of the measurement based on a recalculation of RTP<sub>scCBCT</sub> on planCT (sim) and range differences for the standard recalculation approach (rec). The length of the whiskers was defined as 1.5 times the interquartile ranges.

range difference maps in Fig. 6. Also, the reported median values for the range differences showed high agreement for the two methods for all three patients. The measurement

approach is a measurement-based implementation of the simulation approach using a dosimetry gel phantom. The good agreement of the recalculation and the simulation approach suggests that the measurement and the recalculation approach are equivalent and their results can be compared for analyzing CBCT correction accuracy. Taking the results of all three approaches into account, the median values of the range differences calculated from the dose distributions originating from the CT and the scCBCT are within 0.6 and 2.0 mm. This can be considered acceptable for dose calculation.

As a measurement-based evaluation has not been published so far, only the results of the *recalculation approach* could be compared to existing publications. The presented 3 mm range difference pass rates for the *recalculation approach* of 88.3–97.5% are comparable to the pass rates of 91.9–96.1% reported by Kurz et al.<sup>10</sup> using the same scatter correction method for head and-neck patients. The assumption of negligible anatomical changes, made in the presented study, is not valid for the head and neck cases evaluated in the study of Kurz et al., which were used for comparison. Therefore, their results additionally include an uncertainty due to anatomical variations. To decouple these two sources of uncertainty a different phantom is necessary. In particular, reproducible deformations, as e.g. reported by Niebuhr et al.,<sup>33</sup> need to be combined with the possibility to employ

dosimetry gel for 3D dose measurements. Our study presents a first step to evaluate the impact of image corrections on dose calculation and optimization, but more effort is needed to extend the evaluation for regions with anatomical changes.

High agreement was observed between the film measurement and the simulation (see Fig. 5). Slight deviations were only visible in the caudal region for patients 2 and 3, where the amount of passed bony structure increases. Zhao et al.<sup>20</sup> reported that film dosimetry in near-parallel configuration to the beam is an adequate tool for proton range evaluations with an uncertainty of 0.5 mm. This makes the film measurement ideal for benchmarking the gel results and potential clinical measurement-based quality assurance of in-room CBCT images in adaptive proton therapy.

Although the film measurement matched the simulation better than the gel measurement, the median of the range differences for the gel measurement still showed a deviation to the median of the simulated range differences of less than 1 mm. This is below the registration error estimated for the manual registration and the range determination with gel, both 1 mm.<sup>21</sup> Additionally, the positioning accuracy of the phantom on the treatment table is also in the order of this uncertainty. Although, the good agreement of simulation and measurement indicates that simulation and measurement provide similar information regarding the quality of the scCBCT, an extension of the evaluation for more patients is necessary to answer this question.

A drawback of the presented method is that only one specific beam direction can be tested per gel cylinder or film piece. Although the used phantom allows for easy gel replacement in contrast to the previous phantom design used by Makris et al.<sup>24</sup> and Hillbrand et al.,<sup>21</sup> testing multiple beam direction still requires considerable effort even when the additional costs of separate gel cylinders per beam direction are neglected. While in this work the number of possible beam angles was limited due to the fixed orientation of the film holder, a TPS-based pre-study to evaluate the most relevant beam angle, e.g. in terms of expected deviations, could be considered in further projects using gel alone.

The study design avoided a direct comparison between measurement and calculation in order to remove the error originating from differences between the patient CT and the 3D printed phantom. This made a consideration of a correction of the RSP values unnecessary. Therefore, comprehensive end-to-end results, comparing measured and calculated ranges of the same plan, cannot be provided in the presented workflow.

A further limitation of the study is the use of CBCT image data that were acquired with a photon linear accelerator-mounted CBCT system instead of a dedicated proton CBCT scanner. In addition, other treatment sites have a greater potential for adaptive radiotherapy than the head region, due to more pronounced changes in anatomy. Up to now only a head phantom with the possibility to inserting a gel cylinder was available. For further studies new phantoms for other treatment sites would be desirable.

## 5. CONCLUSIONS

A novel measurement-based evaluation of a scatter corrected CBCT workflow for online adaptive proton radiation therapy was introduced. The evaluation of CBCT image data for proton planning using gel dosimetry showed that the observed range differences agreed well with the expected values from TPS recalculations and optimizations. It is thus an interesting candidate for measurement-based quality assurance of online adaptive proton therapy. The evaluated CBCT correction method seems to be suitable for proton dose calculation. Film measurements provided an additional benchmark in dedicated slices and supported the results obtained with the gel measurements. 98.5% of the range differences observed with the gel measurement agreed with the *simulation approach* within 2 mm. Further studies are needed to evaluate the *measurement-based approach* for more patients and preselected beam directions. A development of a 3D printed phantom for other body regions potentially including anatomical variations would make the method applicable for more treatment sites.

## ACKNOWLEDGMENTS

We thank Emmanouil Zoros, Georgios Kalaitzakis, Kyveli Zourari, Themistoklis Boursianis, and colleagues from RTsafe, P.C. for individualizing the *Prime phantoms*. Katharina Niepel is acknowledged for providing stopping power ratios for the gel.

## CONFLICT OF INTEREST

No potential conflict of interest was reported by the authors.

## FUNDING

This project was supported by the FöFoLe commission of the Medical Faculty of the LMU Munich under grant 994 (F. Kamp) and by the German Research Foundation (DFG) under grant number 399148265 (G. Landry, F. Kamp).

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

\*Present address: Department of Radiation Oncology and Cyberknife Center University Hospital Cologne, University of Cologne, 50937 Cologne, Germany

\*\*Present address: Strahlentherapie Pasing, Munich 81241, Germany

\*\*\*Present address: Strahlentherapie Barmherzige Brüder Krankenhaus, 93049 Regensburg, Germany

<sup>2)</sup> Author to whom correspondence should be addressed. Electronic mail: sebastian.neppi@med.uni-muenchen.de.

## REFERENCES

1. Yan D, Vicini F, Wong J, Martinez A. Adaptive radiation therapy. *Phys Med Biol.* 1997;42(1):123–132.
2. Klüter S. Technical design and concept of a 0.35 T MR-Linac. *Clin Transl. Radiat Oncol.* 2019;18:98–101.
3. Legendijk JJW, Raaymakers BW, Raaijmakers AJE, et al. MRI/linac integration. *Radiother Oncol.* 2008;86:25–29.
4. Yan D, Georg D. Adaptive radiation therapy. *Z Med Phys.* 2018;28:173–174.
5. Lim-Reinders S, Keller BM, Al-Ward S, Sahgal A, Kim A. Online adaptive radiation therapy. *Int J Radiat Oncol Biol Phys.* 2017;99:994–1003.
6. Schulze R, Heil U, Groß D, et al. Artefacts in CBCT: a review. *Dentomaxillofac Radiol.* 2011;40:265–273.
7. Arai K, Kadoya N, Kato T, et al. Feasibility of CBCT-based proton dose calculation using a histogram-matching algorithm in proton beam therapy. *Phys Med.* 2017;33:68–76.
8. Hansen DC, Landry G, Kamp F, et al. ScatterNet: a convolutional neural network for cone-beam CT intensity correction. *Med Phys.* 2018;45:4916–4926.
9. Kurz C, Dedes G, Resch A, et al. Comparing cone-beam CT intensity correction methods for dose recalculation in adaptive intensity-modulated photon and proton therapy for head and neck cancer. *Acta Oncol.* 2015;54:1651–1657.
10. Kurz C, Kamp F, Park Y-K, et al. Investigating deformable image registration and scatter correction for CBCT-based dose calculation in adaptive IMPT. *Med Phys.* 2016;43:5635.
11. Kurz C, Nijhuis R, Reiner M, et al. Feasibility of automated proton therapy plan adaptation for head and neck tumors using cone beam CT images. *Radiat Oncol.* 2016;11:64.
12. Landry G, Dedes G, Zöllner C, et al. Phantom based evaluation of CT to CBCT image registration for proton therapy dose recalculation. *Phys Med Biol.* 2015;60:595–613.
13. Landry G, Hansen D, Kamp F, et al. Corrigendum: Comparing Unet training with three different datasets to correct CBCT images for prostate radiotherapy dose calculations (2019 Phys. Med. Biol 64 035011). *Phys Med Biol.* 2019;64:89501.
14. Landry G, Nijhuis R, Dedes G, et al. Investigating CT to CBCT image registration for head and neck proton therapy as a tool for daily dose recalculation. *Med Phys.* 2015;42:1354–1366.
15. Park Y-K, Sharp GC, Phillips J, Winey BA. Proton dose calculation on scatter-corrected CBCT image: feasibility study for adaptive proton therapy. *Med Phys.* 2015;42:4449–4459.
16. Thummerer A, Zaffino P, Meijers A, et al. Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy. *Phys Med Biol.* 2020;65:95002.
17. Veiga C, Alshaikhi J, Amos R, et al. Cone-beam computed tomography and deformable registration-based “dose of the day” calculations for adaptive proton therapy. *Int J Part Ther.* 2015;2:404–414.
18. Veiga C, Janssens G, Baudier T, et al. A comprehensive evaluation of the accuracy of CBCT and deformable registration based dose calculation in lung proton therapy. *Biomed Phys Eng Express.* 2017;3:15003.
19. Veiga C, Janssens G, Teng C-L, et al. First clinical investigation of cone beam computed tomography and deformable registration for adaptive proton therapy for lung cancer. *Int J Radiat Oncol Biol Phys.* 2016;95:549–559.
20. Zhao L, Das IJ. Gafchromic EBT film dosimetry in proton beams. *Phys Med Biol.* 2010;55:N291–301.
21. Hillbrand M, Landry G, Ebert S, et al. Gel dosimetry for three dimensional proton range measurements in anthropomorphic geometries. *Z Med Phys.* 2019;29:162–172.
22. Niu T, Sun M, Star-Lack J, Gao H, Fan Q, Zhu L. Shading correction for on-board cone-beam CT in radiation therapy using planning MDCT images. *Med Phys.* 2010;37:5395–5406.
23. Zöllner C, Rit S, Kurz C, et al. Decomposing a prior-CT-based cone-beam CT projection correction algorithm into scatter and beam hardening components. *Phys Imaging Radiat Oncol.* 2017;3:49–52.
24. Makris DN, Pappas EP, Zoros E, et al. Characterization of a novel 3D printed patient specific phantom for quality assurance in cranial stereotactic radiosurgery applications. *Phys Med Biol.* 2019;64:105009.
25. Pappas E, Seimenis I, Angelopoulos A, et al. Narrow stereotactic beam profile measurements using N-vinylpyrrolidone based polymer gels and magnetic resonance imaging. *Phys Med Biol.* 2001;46:783–797.
26. van Abbema JK, van Goethem M-J, Mulder J, et al. High accuracy proton relative stopping power measurement. *Nucl Instrum Methods Phys Res B.* 2018;436:99–106.
27. Hillbrand M, Georg D. Assessing a set of optimal user interface parameters for intensity-modulated proton therapy planning. *J Appl Clin Med Phys.* 2010;11:3219.
28. Pappas E, Maris T, Angelopoulos A, et al. A new polymer gel for magnetic resonance imaging (MRI) radiation dosimetry. *Phys Med Biol.* 1999;44:2677–2684.
29. Maris TG, Damilakis J, Sideri L, et al. Assessment of the skeletal status by MR relaxometry techniques of the lumbar spine: comparison with dual X-ray absorptiometry. *Eur J Radiol.* 2004;50:245–256.
30. Micke A, Lewis DF, Yu X. Multichannel film dosimetry with nonuniformity correction. *Med Phys.* 2011;38:2523–2534.
31. Seroul P, Sarrut D. VV: a viewer for the evaluation of 4D image registration. *Midas*, 2008; 1-8
32. Zaffino P, Raudaschl P, Fritscher K, Sharp GC, Spadea MF. Technical Note: plastimatch mabs, an open source tool for automatic image segmentation. *Med Phys.* 2016;43:5155.
33. Niebuhr NI, Johnen W, Echner G, et al. The ADAM-pelvis phantom-an anthropomorphic, deformable and multimodal phantom for MRgRT. *Phys Med Biol.* 2019;64:04NT05.

## Evaluation of proton and photon dose distributions recalculated on 2D and 3D Unet-generated pseudoCTs from T1-weighted MR head scans

Sebastian Neppel<sup>a,b</sup>, Guillaume Landry<sup>a,b</sup>, Christopher Kurz<sup>a,b</sup>, David C. Hansen<sup>c\*</sup>, Ben Hoyle<sup>d</sup>, Sophia Stöcklein<sup>e</sup>, Max Seidensticker<sup>e</sup>, Jochen Weller<sup>d,f</sup>, Claus Belka<sup>a,g</sup>, Katia Parodi<sup>b</sup> and Florian Kamp<sup>a</sup>

<sup>a</sup>Department of Radiation Oncology, University Hospital, LMU Munich, Munich, Germany; <sup>b</sup>Department of Medical Physics, Faculty of Physics, Ludwig-Maximilians-Universität München (LMU Munich), Garching bei München, Germany; <sup>c</sup>Department of Medical Physics, Aarhus University Hospital, Aarhus, Denmark; <sup>d</sup>University Observatory, Faculty of Physics, Ludwig-Maximilians-Universität München (LMU Munich), Munich, Germany; <sup>e</sup>Department of Radiology, University Hospital, LMU Munich, Munich, Germany; <sup>f</sup>Optical and Interpretative Astronomy, Max Planck Institute for Extraterrestrial Physics, Garching bei München, Germany; <sup>g</sup>German Cancer Consortium (DKTK), Partner site Munich, Munich, Germany

### ABSTRACT

**Introduction:** The recent developments of magnetic resonance (MR) based adaptive strategies for photon and, potentially for proton therapy, require a fast and reliable conversion of MR images to X-ray computed tomography (CT) values. CT values are needed for photon and proton dose calculation. The improvement of conversion results employing a 3D deep learning approach is evaluated.

**Material and methods:** A database of 89 T1-weighted MR head scans with about 100 slices each, including rigidly registered CTs, was created. Twenty-eight validation patients were randomly sampled, and four patients were selected for application. The remaining patients were used to train a 2D and a 3D U-shaped convolutional neural network (Unet). A stack size of 32 slices was used for 3D training. For all application cases, volumetric modulated arc therapy photon and single-field uniform dose pencil-beam scanning proton plans at four different gantry angles were optimized for a generic target on the CT and recalculated on 2D and 3D Unet-based pseudoCTs. Mean (absolute) error (MAE/ME) and a gradient sharpness estimate were used to quantify the image quality. Three-dimensional gamma and dose difference analyses were performed for photon (gamma criteria: 1%, 1 mm) and proton dose distributions (gamma criteria: 2%, 2 mm). Range (80% fall off) differences for beam's eye view profiles were evaluated for protons.

**Results:** Training 36 h for 1000 epochs in 3D (6 h for 200 epochs in 2D) yielded a maximum MAE of 147 HU (135 HU) for the application patients. Except for one patient gamma pass rates for photon and proton dose distributions were above 96% for both Unets. Slice discontinuities were reduced for 3D training at the cost of sharpness.

**Conclusions:** Image analysis revealed a slight advantage of 2D Unets compared to 3D Unets. Similar dose calculation performance was reached for the 2D and 3D network.

### ARTICLE HISTORY

Received 1 April 2019  
Accepted 21 May 2019

## Introduction

Magnetic resonance (MR) imaging is an important part of modern radiation therapy treatment planning. In many cases, MR is needed for improved delineation of the target volumes, given the lower soft tissue contrast of X-ray computed tomography (CT) [1,2]. Still, CT images are the basis for treatment planning in terms of determination of electron densities and for estimating relative stopping power ratios to water for photon and proton dose calculation, respectively. Recently, the use of adaptive strategies for photon therapy has increased with the introduction of hybrid machines combining linear accelerators with MR scanners [3,4]. Given that advances in photon radiation therapy are typically


transferred to proton radiation therapy, there has been a growing interest for MR-guided proton therapy research in the last years [5–17]. For online adaptive radiation therapy, replanning strategies have to be fast and reliable. Proton dose calculations are much more sensitive to CT value errors, and therefore require higher conversion accuracy from MR to CT. These converted CTs are often called pseudoCTs or substitute/synthetic CTs. The main challenge for pseudoCT generation is the lack of a physical relationship between MR intensities and electron densities [18].

Several conversion methods have been proposed so far, which were summarized in two recent reviews [18,19]. Deep learning based approaches are the most recent development [20–22]. So far only few approaches directly address proton

**CONTACT** Sebastian Neppel  Sebastian.Neppel@med.uni-muenchen.de  Department of Radiation Oncology, University Hospital, LMU Munich, Munich, Germany

\*Present address: Gradient Software, Aarhus, Denmark.

This article has been republished with minor changes. These changes do not impact the academic content of the article.

 Supplemental data for this article can be accessed [here](#).

dose calculations [23–27]. As none of these approaches is deep learning based, this is the first time a proton dose evaluation is performed on deep learning based pseudoCTs. Deep learning models are highly parallelizable and therefore trained and applied using graphics processing units (GPUs). Due to limitations of the GPU memory, most of the deep learning approaches are trained on two-dimensional (2D) slices. Since the loss functions of 2D models do not account for continuity in the third dimension, slice discontinuities can be observed. Training on three-dimensional (3D) image stacks is expected to achieve a more homogeneous conversion result.

So far, no proton dose calculations were applied to pseudoCTs generated from deep learning models. Our study compared generated pseudoCTs with a U-shaped convolutional neural network for 2D image slices (Unet2D) and a U-shaped convolutional neural network for 3D image stacks (Unet3D) from MR head scans. The comparison was performed by means of image metrics and dose evaluations (gamma and dose difference analysis) for photon and proton dose calculations. For proton dose distributions, additionally, a range analysis was performed.

## Material and methods

### Data description

For this study, CT and MR imaging data of 89 patients with lesions in the brain treated with photon radiation therapy were used. All MR images were acquired with a Magnetom Aera MR scanner (Siemens Healthcare, Germany) with a voxel size below  $1 \times 1 \times 1 \text{ mm}^3$  and a magnetic field strength of 1.5 T using a T1-weighted magnetization-prepared rapid gradient echo (MP-RAGE) sequence (repetition time = 2000 ms, echo time = 3.02 ms, flip angle =  $8^\circ$ ). Sizes of MR images' axial slices varied between  $136 \times 133$  and  $262 \times 262$  pixels. CT scans were from two different scanners, an Acquilion Large Bore CT scanner (Canon Medical Systems, Japan) with a voxel size of  $1.074 \times 1.074 \times (1 \text{ or } 3) \text{ mm}^3$  and a Discovery positron emission tomography (PET)/CT 690 scanner (GE Healthcare, USA) with a voxel size of  $1.273 \times 1.273 \times 2.5 \text{ mm}^3$ . CT axial slices had a standard size of  $512 \times 512$  pixels. CT scans were dedicated for treatment planning and therefore contained patient fixation devices while MR scans were acquired for diagnostic purposes.

The dataset of preprocessed (see [Supporting Information](#) for more details) MR and CT images was split into training (57), validation (28) and application/test (4) patients. The network was trained on the training set, while decisions for the modification of the network, including the stopping point of the training, were made according to the loss values of the validation set. The application data, unseen by the network before, was used to quantify the quality of the generated images and for dose calculations and subsequent analysis.

### Network design

The U-shaped convolutional neural network (Unet) [28] architecture was initially designed for Hounsfield unit (HU)

correction of cone beam computed tomography images [29,30] and was slightly adapted for packages of 32 slices (Unet3D) (see [Supporting Information](#)).

### Dose calculation

To evaluate dosimetric properties of the generated CTs, a single-arc volumetric modulated arc therapy (VMAT) plan and single-field uniform dose (SFUD) pencil-beam scanning (PBS) plans from four different gantry angles ( $45^\circ$ ,  $135^\circ$ ,  $225^\circ$ ,  $315^\circ$ ) were optimized in a research version of the clinical treatment planning software RayStation (version 6.99, RaySearch Laboratories AB, Sweden). Photon dose calculations were performed using a collapsed-cone algorithm with a beam model for a Synergy linear accelerator (Elekta, Sweden) and a dose grid size of  $3 \times 3 \times 3 \text{ mm}^3$  on the CT. The resulting plans were recalculated on the corresponding pseudoCTs generated by the two Unets. In order to make range analysis in axial slices for protons more accurate a dose grid size of  $1 \times 1 \times 3 \text{ mm}^3$  was applied for the IBA Dedicated beam model. A median dose of 60 Gy in 30 fractions was aimed for a generic target volume in all plans. For all optimized dose distributions, the target volume fraction receiving more than 95% of the prescribed dose ( $V_{95\%}$ ) was above 99%.

### Data evaluation

The quality of the generated pseudoCTs was assessed by calculating the mean and absolute error (MAE/ME), both given in HUs within the outer contour of the patient. A sharpness estimate was determined by a gradient-based measure (see [Supporting Information](#)).

Dose distributions of the CTs were the references for dose distributions on the pseudoCTs. For VMAT plans the dose was evaluated by means of a 3D gamma analysis [31] with a 1 mm distance and a global 1% dose criterion ( $\Gamma_{1\%,1\text{mm}}$ ). Additionally, a dose difference analysis with 1% passing criterion ( $DD_{1\%}$ ) was applied. For photon doses, an evaluation dose threshold of 20% was used. For proton plans besides  $\Gamma_{2\%,2\text{mm}}$  and  $DD_{2\%}$ , both with a dose threshold of 50%, a range difference analysis of the beam's eyes profiles was performed. The application of the gamma and dose difference analysis incorporated an interpolation of the dose to a 1 mm dose grid. The range was defined as the distance between the 80% dose fall-off and the respective proximal patient outline in beam direction. Pass rates for the range differences exceeding 2 mm ( $RD_{2\text{mm}}$ ) and 3 mm ( $RD_{3\text{mm}}$ ) were calculated.

## Results

### Training

From 1000 epochs of 3D and 200 epochs of 2D training, epochs 908 and 158 were selected for Unet3D and Unet2D, respectively (see [Supplementary material](#) for more information). Training duration was about 36 h for Unet3D and 6 h



for Unet2D. The conversion time per patient ( $\approx 100$  slices) was lower than 2 s for Unet2D and 12 s for Unet3D.

### Image analysis

Figure 1 shows a sagittal slice of the pseudoCT of one of the application patients for both Unets, the respective CT scan and the difference plots (generated pseudoCT - CT) with the corresponding MR scan. The slice discontinuities which are visible at tissue-air-interfaces for Unet2D, did not appear for Unet3D, but the output of Unet3D compared to Unet2D is clearly blurrier. This visual impression is supported by the sharpness estimate, which yielded for the four application patients (95%, 96%, 99%, 95%) and (85%, 88%, 98%, 93%) for Unet2D and Unet3D, respectively. The increased blur of the Unet3D results also reduced the capability to capture fine bony structures. Large HU differences appeared mainly in the regions next to bony structures, but positional shifts were not observed (Figure 1).

Figure 2 presents the MAE and ME in HU for both Unets for the whole patient set. The average MAE values over all patients of the training set given with one standard deviation were lower for Unet2D ( $55 \pm 10$  HU) compared to Unet3D ( $90 \pm 20$  HU). Although this trend was also visible for the validation set, the mean MAE values for the validation set agreed within one standard deviation with  $116 \pm 26$  HU and  $137 \pm 32$  HU for Unet2D and Unet3D, respectively. For the application set, MAE values ranged from 82 HU to 135 HU for Unet2D and from 82 HU to 147 HU for Unet3D.

In terms of ME, Unet2D exhibited no clear systematic ME deviation for the training patients and values up to  $\pm 15$  HU were observed. In contrast, Unet3D showed mainly positive deviations up to 40 HU for the training set. Average ME values over the training ( $-1 \pm 4$  HU) and validation patient group ( $2 \pm 30$  HU) were close to zero for Unet2D. Unet3D showed a higher spread in ME values yielding  $11 \pm 9$  HU and  $31 \pm 55$  HU for training and validation set, respectively.

### Dose analysis

Evaluation of the VMAT plans recalculated on the pseudoCTs of Unet2D and Unet3D with  $\Gamma_{1\%,1\text{mm}}$  and  $DD_{1\%}$  yielded for all application patients pass rates better than 95%. For 2% dose difference evaluation, the pass rates were above 98%. Given these high values no further VMAT results are presented.

For protons, an axial view of the dose distributions for the 315° SFUD plan of patient 50 and patient 60 (application set) is illustrated in Figure 3. The dose distribution of the plan on the CT is shown together with the recalculated dose distributions on the pseudoCTs generated by Unet2D and Unet3D. Additionally, differences between the CT and MR patient outlines were plotted on top of the respective MR slices. Considerable differences between the patient outlines and between the target volume and the 95% isodose line were observed for patient 60.

Table 1 presents  $\Gamma_{2\%,2\text{mm}}$ ,  $DD_{2\%}$ ,  $RD_{3\text{mm}}$  and  $RD_{2\text{mm}}$  of the dose distributions for the proton plans of the application

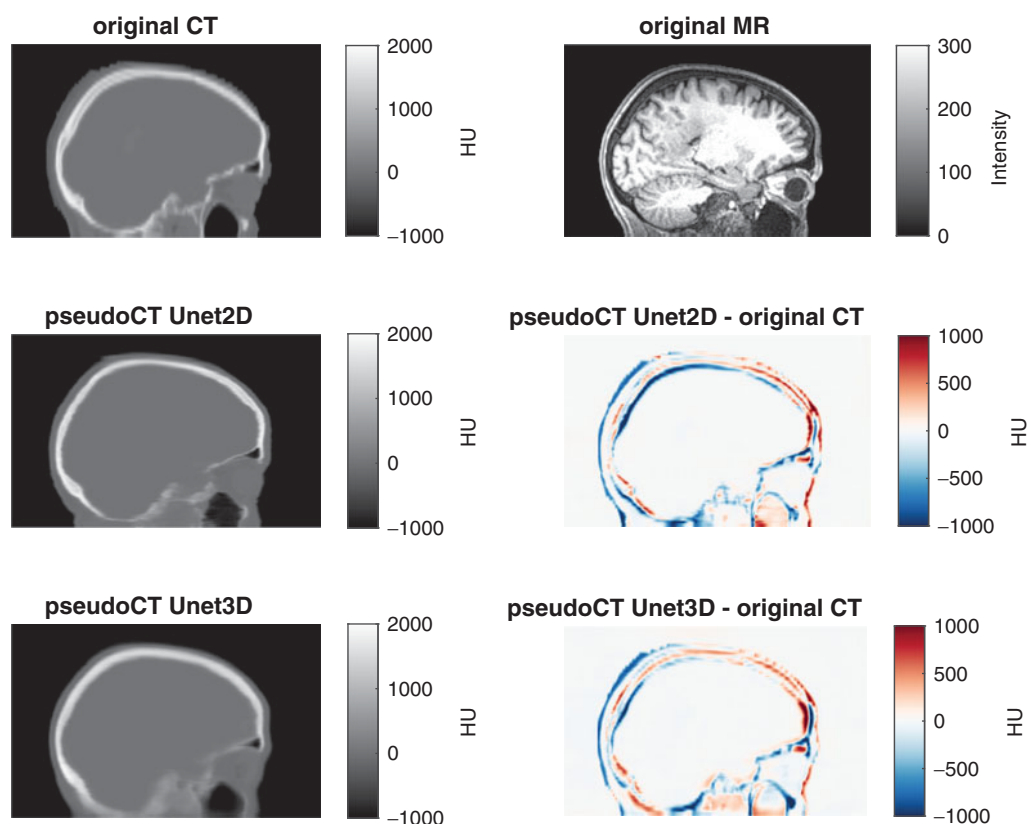


Figure 1. Sagittal views for patient 56 (application set) of the CT (upper left), the pseudoCT of Unet2D (middle left), the pseudoCT of Unet3D (lower left) and the respective MR (upper right). The difference plots pseudoCT Unet2D - original CT (middle right) and pseudoCT Unet3D - original CT (lower right) are also shown.

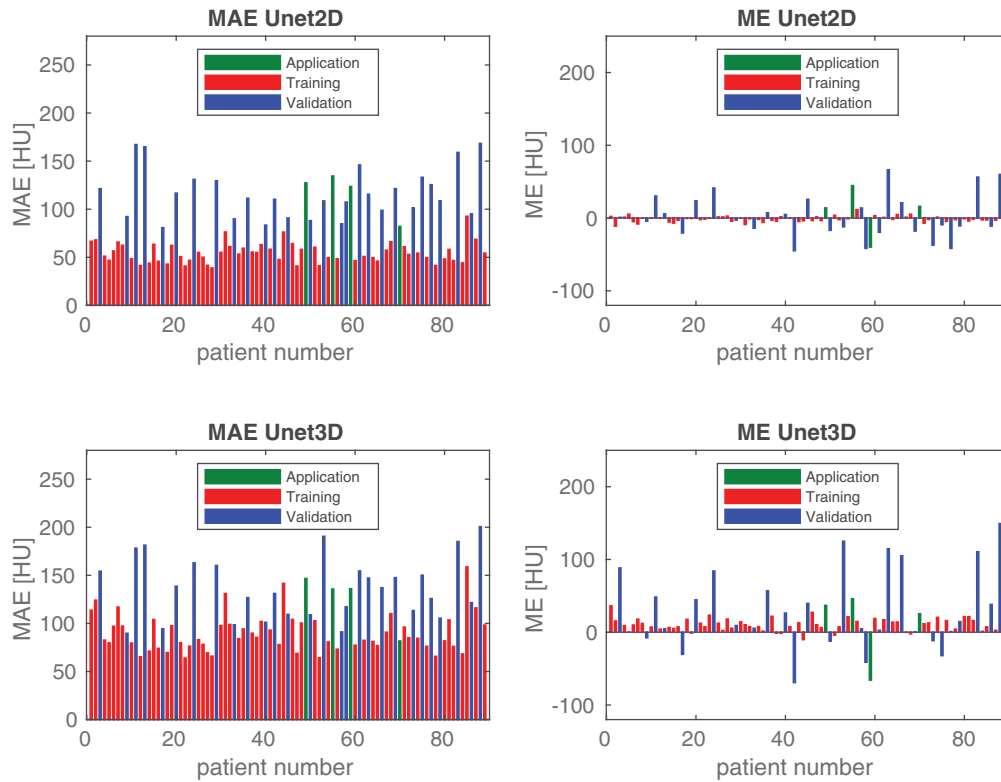


Figure 2. The MAE (left column) and ME (right column) values plotted for the three patient groups (training, validation and application) for Unet2D and Unet3D.

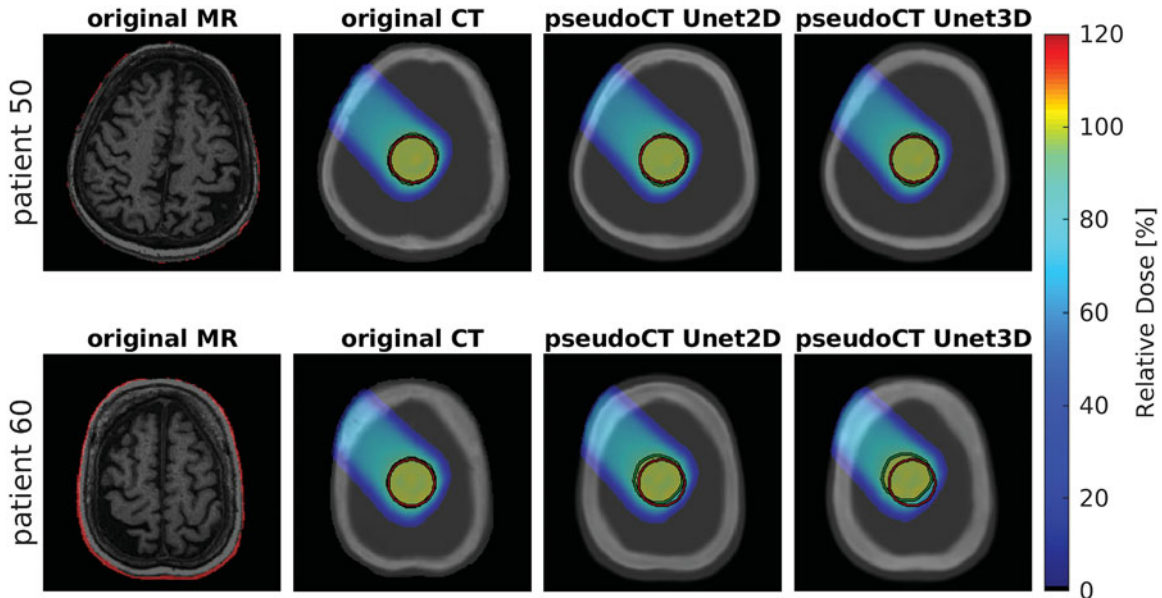


Figure 3. Axial view of the SFUD proton dose distributions for gantry angle  $315^\circ$  of patient 50 and patient 60 (application set) for the CT, the pseudoCT generated by Unet2D and the pseudoCT generated by Unet3D. The generic target is indicated in red and the 95% isodose in green. The dose values were normalized to the prescribed dose. The respective MR slices are shown to the left with the differences between CT and MR patient outlines marked in red.

patients and the corresponding mean values over all proton plans and application patients.  $\Gamma_{2\%,2\text{mm}}$  for the dose distributions recalculated on the pseudoCT of Unet2D and Unet3D were all above 97%, except for patient 60 which had the lowest pass rate for the plan from gantry angle  $225^\circ$  yielding 91.7% and 89.3% for Unet2D and Unet3D, respectively.  $DD_{2\%}$  yielded around 5% lower pass rates than the gamma evaluation. Low pass rates obtained for patient 60 by gamma

and dose difference calculations were even more pronounced for the range differences. Range differences of patient 60 were above 2 mm for 44% of the profiles for gantry angle  $315^\circ$ , while less than 10.5% of the profiles of the remaining application patients had range differences above 2 mm. Three millimeter range differences were observed for less than 5% of the profiles disregarding patient 60 for both Unets.

**Table 1.** Pass rates in percent for 2%/2 mm gamma ( $\Gamma_{2\%,2\text{mm}}$ ), 2% dose difference ( $DD_{2\%}$ ), 2 mm range ( $RD_{2\text{mm}}$ ) and 3 mm range difference analysis ( $RD_{3\text{mm}}$ ) for both Unets (2D and 3D) and all proton dose distributions of the application patient set.

Type	Patient	45°		135°		225°		315°		$\mu \pm \sigma$	
		2D	3D	2D	3D	2D	3D	2D	3D	2D	3D
$\Gamma_{2\%,2\text{mm}}$	50	99.4	99.2	99.3	99.1	98.0	98.3	99.2	98.0	98 ± 2	97 ± 3
	56	98.5	97.2	98.8	99.1	98.4	99.1	98.2	97.1		
	60	98.4	96.1	95.0	91.4	91.7	89.3	93.6	90.8		
	71	99.5	99.3	99.6	99.4	99.1	99.0	98.9	99.2		
$DD_{2\%}$	50	95.6	95.8	95.4	93.2	92.9	93.6	95.5	92.1	93 ± 3	92 ± 4
	56	93.2	91.4	92.7	92.5	93.1	93.6	91.5	90.6		
	60	93.2	89.5	90.4	87.1	87.3	85.5	87.5	85.4		
	71	97.2	96.8	96.8	96.0	94.7	94.4	95.0	96.4		
$RD_{2\text{mm}}$	50	98.2	94.2	95.5	94.8	96.1	95.5	98.2	94.2	93 ± 4	90 ± 13
	56	92.0	94.9	97.4	98.4	90.7	98.6	88.7	97.0		
	60	93.1	85.1	95.3	76.4	84.6	66.0	93.5	56.0		
	71	97.2	98.2	89.6	97.9	91.5	98.1	92.8	99.0		
$RD_{3\text{mm}}$	50	100	100	99.1	98.8	99.5	99.8	98.8	96.9	98 ± 2	96 ± 7
	56	96.1	98.5	99.6	99.9	99.3	99.8	97.2	99.8		
	60	97.5	92.8	99.5	87.0	95.7	79.0	97.7	84.1		
	71	99.5	99.8	94.8	99.4	96.0	100	98.7	99.9		

The mean and one standard deviation over all application patients and proton plans ( $\mu \pm \sigma$ ) is given for all evaluation types in the last two rows.

## Discussion

PseudoCTs generated from a 2D and a 3D Unet were applied to proton therapy for the first time. Both approaches were compared in terms of their photon and proton dose calculation accuracy. The 3D patch size of  $256 \times 256 \times 32$  allowed to feed complete slice stacks with a good image resolution into the network. This helped to reduce slice discontinuities observed for Unet2D (see Figure 1). Other groups reported 3D patch sizes of e.g.,  $32 \times 32 \times 12$  [32] not allowing to feed in complete 2D image slices of a patient. In terms of sharpness the resulting pseudoCTs of Unet3D were blurrier than those of Unet2D, potentially due to a reduced training sample size between 2D and 3D. This could possibly be improved by increasing the number of patient datasets.

Short application times are crucial for the acceptance of the method for online adaptive replanning strategies. Our reported application times of about 10 s for Unet3D are comparable to those reported in literature [20,21]. For Unet2D, application time was more than a factor of five lower. Using only the central slice of each image stack for application increased the Unet3D application time because of a large overlap of the stacks. The typically much longer training times of several hours up to days are not an issue as they only have to be performed once. Since the method only relies on one MR sequence, the time the patient lies on the table is reduced compared to strategies that need two or more different MR sequences [33,34].

ME values of the training set were smaller and less biased for Unet2D compared to Unet3D (see Figure 2). For Unet2D, no bias was visible neither for the validation nor for the application set. This means that for Unet3D average HU values are slightly higher compared to the CT. This trend is less prominent for the validation set.

MAE values were much smaller for training than for validation and application set for both Unets. This could be an indicator that the networks were overfitting the training data. This assumption is not supported by the loss curves (see Supporting Information, Figure S2), since no increase in

the validation loss curves was observed. Compared to the range of MAE values achieved by other publications for brain pseudoCT generation our calculated MAE maxima for the two Unets over all application patients are in the range reported for mean MAE values in the head region (85–230 HU) by the review of Edmund and Nyholm [18].

Nonetheless, dose evaluations for photons yielded very good pass rates for both Unets. Proton dose calculation yielded good pass rates for three out of four patients. The gamma passing rates for the proton plans were above 89.3% and hence in the range reported by Pillegi et al. [25], although the values are not directly comparable as they only perform the gamma analysis within the planning target volume (PTV) in the ten central slices of the dose cube. The exception was patient 60 who had generally lower passing rates. Range differences showed the same behavior. Low pass rates for patient 60 can be explained by the information shown in Figure 3. A difference in the patient outline between CT and MR and an overestimation of the bone in the upper part of the skull cause comparably large range differences. A mismatch of the rigid registration was excluded by visual inspection of a clinical expert. Changes in the patient outline might be caused by the mask and the pillow, which were not present in the MR. Since the differences mainly appear in the upper part of the skull, where the size of the patient outline strongly changes from slice to slice, the interpolation from CT (3 mm) to MR slice distance (1 mm) might be another influencing factor.

For proton therapy, a direct conversion from MR intensity to relative stopping power ratio would be favorable. In this work, a more clinically oriented approach was chosen. By converting the MR intensities to CT values instead of relative stopping power ratios, the same standard conversion routine could be used.

In conclusion, our Unet-based approach yielded comparable results as previously reported pseudoCT generation methods for proton dose calculations in the brain at considerably lower application times in the 2D case. Image analysis revealed a slight advantage of Unet2D, but for photon and proton dose calculations similar accuracy was achieved.

## Acknowledgments

We thank Erik Traneus and colleagues from Raysearch Laboratories, as well as Marco Pinto, for support on the research version of the treatment planning system used in this study.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

The P6000 used in this work was donated by the Nvidia Corporation. Florian Kamp was supported by the FöFoLe commission of the Medical Faculty of the LMU Munich under grant 994. Christopher Kurz was supported by the German Cancer Aid (Deutsche Krebshilfe). This work was supported by the German Research Foundation (DFG) Cluster of Excellence Munich Center for Advanced Photonics (MAP).

## References

- [1] Khoo VS, Joon DL. New developments in MRI for target volume delineation in radiotherapy. *Br J Radiol.* 2006;79 Spec No 1: S2–S15.
- [2] Dirix P, Haustermans K, Vandecaveye V. The value of magnetic resonance imaging for radiotherapy planning. *Semin Radiat Oncol.* 2014;24:151–159.
- [3] Mutic S, Dempsey JF. The ViewRay system: magnetic resonance-guided and controlled radiotherapy. *Semin Radiat Oncol.* 2014;24: 196–199.
- [4] Raaymakers BW, Lagendijk JJW, Overweg J, et al. Integrating a 1.5 T MRI scanner with a 6 MV accelerator: proof of concept. *Phys Med Biol.* 2009;54:N229–N237.
- [5] Raaymakers BW, Raaijmakers AJE, Lagendijk J. Feasibility of MRI guided proton therapy: magnetic field dose effects. *Phys Med Biol.* 2008;53:5615–5622.
- [6] Crijns SPM, Raaymakers BW, Lagendijk J. Real-time correction of magnetic field inhomogeneity-induced image distortions for MRI-guided conventional and proton radiotherapy. *Phys Med Biol.* 2011;56:289–297.
- [7] Schippers JM, Lomax AJ. Emerging technologies in proton therapy. *Acta Oncol.* 2011;50:838–850.
- [8] Moteabbed M, Schuemann J, Paganetti H. Dosimetric feasibility of real-time MRI-guided proton therapy. *Med Phys.* 2014;41: 111713.
- [9] Oborn BM, Dowdell S, Metcalfe PE, et al. Proton beam deflection in MRI fields: implications for MRI-guided proton therapy. *Med Phys.* 2015;42:2113–2124.
- [10] Kurz C, Landry G, Resch AF, et al. A Monte-Carlo study to assess the effect of 1.5 T magnetic fields on the overall robustness of pencil-beam scanning proton radiotherapy plans for prostate cancer. *Phys Med Biol.* 2017;62:8470–8482.
- [11] Oborn BM, Dowdell S, Metcalfe PE, et al. Future of medical physics: real-time MRI-guided proton therapy. *Med Phys.* 2017;44: e77–e90.
- [12] Schellhammer SM, Hoffmann AL. Prediction and compensation of magnetic beam deflection in MR-integrated proton therapy: a method optimized regarding accuracy, versatility and speed. *Phys Med Biol.* 2017;62:1548–1564.
- [13] Shao W, Tang X, Bai Y, et al. Modulation of lateral positions of Bragg peaks via magnetic fields inside cancer patients: Toward magnetic field modulated proton therapy. *Med Phys.* 2017;44: 5325–5338.
- [14] Paganelli C, Whelan B, Peroni M, et al. MRI-guidance for motion management in external beam radiotherapy: current status and future challenges. *Phys Med Biol.* 2018;63:22TR03.
- [15] Schellhammer SM, Gantz S, Lühr A, et al. Technical note: experimental verification of magnetic field-induced beam deflection and Bragg peak displacement for MR-integrated proton therapy. *Med Phys.* 2018;45:3429–3434.
- [16] Schellhammer SM, Hoffmann AL, Gantz S, et al. Integrating a low-field open MR scanner with a static proton research beam line: proof of concept. *Phys Med Biol.* 2018;63:23LT01.
- [17] Santos DM, Wachowicz K, Burke B, et al. Proton beam behavior in a parallel configured MRI-proton therapy hybrid: effects of time-varying gradient magnetic fields. *Med Phys.* 2019;46:822–838.
- [18] Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiat Oncol.* 2017;12:28.
- [19] Johnstone E, Wyatt JJ, Henry AM, et al. Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging-only radiation therapy. *Int J Radiat Oncol Biol Phys.* 2018;100:199–217.
- [20] Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys.* 2017;44:1408–1419.
- [21] Wolterink JM, Dinkla AM, Savenije MHF, et al. Deep MR to CT synthesis using unpaired data. Tsafaris SA, Gooya A, Frangi AF, Prince JL, editors. *Simulation and Synthesis in Medical Imaging: Second International Workshop, SASHIMI 2017, held in conjunction with MICCAI 2017, Québec City, QC, Canada, September 10, 2017: proceedings.* Cham, Switzerland: Springer; 2017. p. 14–23.
- [22] Dinkla AM, Wolterink JM, Maspero M, et al. MR-only brain radiation therapy: dosimetric evaluation of synthetic CTs generated by a dilated convolutional neural network. *Int J Radiat Oncol Biol Phys.* 2018;102:801–812.
- [23] Rank CM, Hünemohr N, Nagel AM, et al. MRI-based simulation of treatment plans for ion radiotherapy in the brain region. *Radiat Oncol.* 2013;109:414–418.
- [24] Koivula L, Wee L, Korhonen J. Feasibility of MRI-only treatment planning for proton therapy in brain and prostate cancers: dose calculation accuracy in substitute CT images. *Med Phys.* 2016;43: 4634.
- [25] Pileggi G, Speier C, Sharp GC, et al. Proton range shift analysis on brain pseudo-CT generated from T1 and T2 MR. *Acta Oncol.* 2018;57:1521–1531.
- [26] Maspero M, van den Berg CAT, Landry G, et al. Feasibility of MR-only proton dose calculations for prostate cancer radiotherapy using a commercial pseudo-CT generation method. *Phys Med Biol.* 2017;62:9159–9176.
- [27] Guerreiro F, Koivula L, Seravalli E, et al. Feasibility of MRI-only photon and proton dose calculations for pediatric patients with abdominal tumors. *Phys Med Biol.* 2019;64:055010.
- [28] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Medical image computing and computer-assisted intervention - MICCAI. 1st ed.* Cham: Springer International Publishing; Imprint: Springer; 2015. p. 234–241.
- [29] Hansen DC, Landry G, Kamp F, et al. ScatterNet: a convolutional neural network for cone-beam CT intensity correction. *Med Phys.* 2018;45:4916–4926.
- [30] Landry G, Hansen D, Kamp F, et al. Comparing U-net training with three different datasets to correct CBCT images for prostate radiotherapy dose calculations. *Phys Med Biol.* 2019;64:035011.
- [31] Low DA, Harms WB, Mutic S, et al. A technique for the quantitative evaluation of dose distributions. *Med Phys.* 1998;25:656–661.
- [32] Nie D, Cao X, Gao Y, et al. Estimating CT image from MRI data using 3D fully convolutional networks. *Deep learning and data labeling for medical applications.* Cham: Springer International Publishing; 2016.
- [33] Kapanen M, Tenhunen M. T1/T2\*-weighted MRI provides clinically relevant pseudo-CT density data for the pelvic bones in MRI-only based radiotherapy treatment planning. *Acta Oncol.* 2013;52: 612–618.
- [34] Korhonen J, Kapanen M, Keyriläinen J, et al. Influence of MRI-based bone outline definition errors on external radiotherapy dose calculation accuracy in heterogeneous pseudo-CT images of prostate cancer patients. *Acta Oncol.* 2014;53:1100–1106.



# Real-time 4DMRI-based internal target volume definition for moving lung tumors

Moritz Rabe<sup>a)</sup> and Christian Thieke

*Department of Radiation Oncology, University Hospital, LMU Munich, Munich 81377, Germany*

Mathias Düsberg

*Department of Radiation Oncology, Klinikum rechts der Isar, Technical University, Munich 81675, Germany*

Sebastian Neppel, Sabine Gerum, and Michael Reiner

*Department of Radiation Oncology, University Hospital, LMU Munich, Munich 81377, Germany*

Nils Henrik Nicolay

*Medical Center, University of Freiburg, Freiburg 79106, Germany*

Heinz-Peter Schlemmer

*Department of Radiology, German Cancer Research Center, Heidelberg 69120, Germany*

Jürgen Debus

*Department of Radiation Oncology, University Hospital of Heidelberg, Heidelberg 69120, Germany*

*Heidelberg Institute of Radiation Oncology (HIRO), Heidelberg 69120, Germany*

Julien Dinkel

*Department of Radiology, University Hospital, LMU Munich, Munich 81377, Germany*

Guillaume Landry

*Department of Radiation Oncology, University Hospital, LMU Munich, Munich 81377, Germany*

*Department of Medical Physics, Ludwig-Maximilians-Universität München (LMU Munich), Garching 85748, Germany*

Katia Parodi

*Department of Medical Physics, Ludwig-Maximilians-Universität München (LMU Munich), Garching 85748, Germany*

Claus Belka

*Department of Radiation Oncology, University Hospital, LMU Munich, Munich 81377, Germany*

*German Cancer Consortium (DKTK), Munich 81377, Germany*

Christopher Kurz\*

*Department of Radiation Oncology, University Hospital, LMU Munich, Munich 81377, Germany*

*Department of Medical Physics, Ludwig-Maximilians-Universität München (LMU Munich), Garching 85748, Germany*

Florian Kamp\*

*Department of Radiation Oncology, University Hospital, LMU Munich, Munich 81377, Germany*

(Received 2 August 2019; revised 20 December 2019; accepted for publication 7 January 2020; published 10 February 2020)

**Purpose:** In photon radiotherapy, respiratory-induced target motion can be accounted for by internal target volumes (ITV) or mid-ventilation target volumes (midV) defined on the basis of four-dimensional computed tomography (4D-CT). Intrinsic limitations of these approaches can result in target volumes that are not representative for the gross tumor volume (GTV) motion over the course of treatment. To address these limitations, we propose a novel patient-specific ITV definition method based on real-time 4D magnetic resonance imaging (rt-4DMRI).

**Methods:** Three lung cancer patients underwent weekly rt-4DMRI scans. A total of 24 datasets were included in this retrospective study. The GTV was contoured on breath-hold MR images and propagated to all rt-4DMRI images by deformable image registration. Different targets were created for the first (reference) imaging sessions: ITVs encompassing all GTV positions over the complete (ITV<sup>80s</sup>) or partial acquisition time (ITV<sup>10s</sup>), ITVs including only voxels with a GTV probability-of-presence (POP) of at least 5% (ITV<sup>5%</sup>) or 10% (ITV<sup>10%</sup>), and the mid-ventilation GTV position. Reference planning target volumes (PTV<sub>r</sub>) were created by adding margins around the ITVs and midV target volumes. The geometrical overlap of the PTV<sub>r</sub> with ITV<sub>n</sub><sup>5%</sup> from the six to eight subsequent imaging sessions on days *n* was quantified in terms of the Dice similarity coefficient (DSC), sensitivity [SE: (PTV<sub>r</sub> ∩ ITV<sub>n</sub><sup>5%</sup>)/ITV<sub>n</sub><sup>5%</sup>] and precision [PRE: (PTV<sub>r</sub> ∩ ITV<sub>n</sub><sup>5%</sup>)/PTV<sub>r</sub>] as surrogates for target coverage and normal tissue sparing.

**Results:** Patient-specific analysis yielded a high variance of the overlap values of PTV<sub>r</sub><sup>10s</sup>, when different periods within the reference imaging session were sampled. The mid-ventilation-based



PTVs were smaller than the ITV-based PTVs. While the SE was high for patients with small breathing pattern variations, changes of the median breathing amplitudes in different imaging sessions led to inferior SE values for the mid-ventilation PTV for one patient. In contrast,  $PTV_r^{5\%}$  and  $PTV_r^{10\%}$  showed higher SE values with a higher robustness against interfractional changes, at the cost of larger target volumes.

**Conclusions:** The results indicate that rt-4DMRI could be valuable for the definition of target volumes based on the GTV POP to achieve a higher robustness against interfractional changes than feasible with today's 4D-CT-based target definition concepts. © 2020 The Authors. *Medical Physics* published by Wiley Periodicals, Inc. on behalf of American Association of Physicists in Medicine. [https://doi.org/10.1002/mp.14023]

Key words: 4DMRI, interfractional changes, ITV, lung tumor, mid-ventilation, motion management

## 1. INTRODUCTION

In high-precision radiotherapy (RT) of lung tumors, target motion due to respiration remains a predominant challenge.<sup>1</sup> This motion can be substantial, is patient-specific, difficult to predict, can be irregular and change from one day to another.<sup>2–8</sup> Intrafractional motion, defined as any motion induced by physiological processes occurring within a treatment session such as breathing, and interfractional changes occurring between treatment sessions need to be accounted for in the planning and delivery process.<sup>9</sup>

To address intrafractional changes, passive and active motion management techniques have been developed over the last decades.<sup>10</sup> Passive methods include motion-encompassing margins and abdominal compression. Active methods include active breathing control (ABC), breath-hold techniques, gating, and tracking. While active approaches have a higher potential in reducing the integral dose to the patient, they can be invasive when fiducial markers are implanted, complex, costly, time-consuming, require specialized equipment, or are still in the research phase. The clinical benefit of many of these methods remains to be proven.<sup>3,4</sup> For these reasons, passive motion-management (PMM) techniques are still primarily used clinically, in particular for conventionally fractionated RT.<sup>1,10</sup> The use of internal target volumes (ITV) as a motion-encompassing method is described in Report 83 of the International Commission on Radiation Units and Measurements (ICRU).<sup>11</sup> Assessment of the range of motion by four-dimensional computed tomography (4D-CT) imaging has become the clinical standard-of-care.<sup>4,12</sup> The ITV is ideally obtained from the union of all gross tumor volumes (GTVs) delineated on the datasets at the different breathing phases.<sup>7</sup> It ideally includes all possible positions of the GTV throughout the course of treatment.<sup>13</sup> The ITV is then expanded by margins to account for interfractional changes and patient setup uncertainties to create the planning target volume (PTV). An alternative PMM technique is based on the definition of the mid-position<sup>14</sup> or mid-ventilation target volume (midV),<sup>15</sup> where the average position of the GTV is reconstructed from the 4D-CT and motion-dependent anisotropic margins are added to account for the respiration-induced target motion. The resulting PTVs are typically smaller than corresponding ITVs.<sup>4,16</sup> Although the midV concept has the potential to

reduce the integral dose to the lungs, its clinical implementation is typically limited to academic RT centers.<sup>17</sup>

It is questionable whether the characterization of respiratory motion and the derivation of a corresponding PTV based on a single pretreatment 4D-CT is representative and adequate for treatment planning.<sup>13,18–21</sup> 4D-CT images are averaged over only a few breathing cycles and the target volume defined based on these images is therefore subject to random uncertainties,<sup>22</sup> since the breathing of the patient during this short acquisition time might not be representative for the patient's breathing pattern during treatment. Clinically relevant interfractional anatomical changes such as tumor shrinkage, weight loss or normal tissue alterations like pleural effusion, and onset or resolution of atelectasis are frequently observed during RT of lung tumors.<sup>5,9,10</sup> The patient's breathing pattern can change due to psychological factors, such as an increase of relaxation of the patient over the course of treatment.<sup>23</sup> These changes are in general not predictable but introduce systematic errors in the treatment that can compromise the quality of the RT treatment due to reduced target coverage or additional dose to organs at risk.<sup>1,24</sup>

Online-adaptive magnetic resonance imaging (MRI)-guided RT may deliver highly conformal doses to the tumor by adjusting the treatment plan just before the treatment session if interfractional changes occurred.<sup>25</sup> The equipment needed for this method is, however, not widespread.<sup>26</sup> Today, the majority of clinics use daily volumetric imaging with three-dimensional cone-beam CT (CBCT) for positioning in image-guided RT (IGRT) of lung cancer<sup>27</sup> through which relevant interfractional anatomical changes can be detected. Four-dimensional CBCT imaging can improve the accuracy of patient positioning compared to 3D-CBCT<sup>28</sup> and enables a motion assessment directly before treatment. This motion assessment is subject to random uncertainties related to the averaging of motion due to the reconstruction of a single breathing cycle from projections from several breathing cycles with potential inter-cyclic variations.<sup>3</sup> Therefore, the question whether the 4D-CT-based ITV used for treatment planning is still adequately representing the target motion cannot be answered.

The introduction of MRI in the RT workflow<sup>29</sup> has spurred research that could potentially contribute to the solution of this problem. The movement of lung tumors has been investigated with two-dimensional (2D) cine-MRI, which

enables the repeated acquisition of temporally resolved images of arbitrary duration with high soft tissue contrast without delivering dose to the patient.<sup>5,6</sup> Several studies have investigated intrafractional and interfractional changes with 2D cine-MRI.<sup>6,30–35</sup> Cai et al.<sup>36</sup> showed that the characterization of lung tumor motion with temporally resolved MRI is more representative than with 4D-CT, mainly due to improved statistics through longer acquisition times.<sup>22</sup> Hence, temporally resolved MRI can help to refine the ITV definition for treatment planning.

When 2D cine-MR images are used, the information about the target motion is limited to 2D planes and only indirect inferences of the out-of-plane motion of tumor and surrounding tissues can be drawn.<sup>37</sup> The extension to four-dimensional MRI (4DMRI) is therefore an active field of research,<sup>38,39</sup> through which the 3D motion of the target and surrounding tissues including translations, rotations and deformations can be directly captured. As pointed out in a recent review on 4DMRI,<sup>39</sup> research is mainly focused on respiratory-correlated 4DMRI (rc-4DMRI) opposed to real-time 4DMRI (rt-4DMRI).<sup>39–42</sup> In contrast to 4D-CT imaging and rc-4DMRI, no retrospective sorting of projections or 2D images from different breathing cycles based on a surrogate is needed for rt-4DMRI. Fast 3D gradient echo (GRE) or steady-state free precession (SSFP) sequences using parallel imaging techniques and echo sharing are usually used for rt-4DMRI.<sup>42</sup> The in-plane resolution is 3–4 mm and temporal resolution is typically limited to 2 volumes per second, depending on the spatial resolution.<sup>39</sup> rt-4DMRI is not routinely used clinically today, but a more widespread use due to the technological advances in this area is expected in the near future.<sup>39</sup>

The purpose of this proof-of-concept study is to demonstrate how rt-4DMRI could be used to reduce the random and systematic uncertainties associated with today's 4D-CT-based ITV definition approach. We describe how a 4DMRI-based ITV can be defined based on the probability-of-presence (POP) of the GTV to reduce random uncertainties. Additional PTV margins are added to reduce systematic uncertainties and prospectively account for potential interfractional changes. The new ITV definition concept is evaluated by analyzing the geometrical overlap of the obtained PTV with ITVs based on the GTV motion on different days. The method is compared to today's PMM concepts including PTVs that mimic a 4D-CT-based ITV definition and the mid-ventilation approach. We show how systematic errors due to interfractional changes could be quantified with regular 4D-MR imaging by metrics that are correlated to dosimetric quantities such as target coverage and normal tissue sparing. The potential integration of the proposed novel ITV definition into clinical practice is outlined and its limitations are discussed.

## 2. MATERIALS AND METHODS

### 2.A. Patient data and imaging protocols

Three patients with tumors in the right lung were included in this retrospective proof-of-concept study. The patients

underwent regular MR imaging in treatment position with a 1.5 T scanner (Siemens Avanto) with 7–9 imaging sessions distributed over 11–12 weeks. A total of 24 datasets were acquired for all patients accumulated. A 3D image in breath-hold (3DMRI) was acquired with a balanced steady-state free precession (bSSFP) sequence (TrueFISP; axial slices; slice thickness: 4–5 mm; in-plane resolution:  $0.88 \times 0.88 \text{ mm}^2$ ; TR/TE: 380/1.16 ms; flip angle:  $63^\circ$ ; field-of-view (FOV):  $45 \times 45 \times 24 \text{ cm}^3$ ; receiver bandwidth: 1030 Hz/px) using parallel imaging (GRAPPA) at the beginning of each session. The acquisition time for each axial slice was 400 ms, resulting in a total acquisition time of 20–24 s for the whole 3D volume. An rt-4DMRI dataset (4DMRI) was subsequently acquired. A total of 157 3D volumes were collected over a period of 80 s with a temporal resolution of 500 ms and the patient breathing freely (TWIST; coronal slices; slice thickness: 10 mm; in-plane resolution:  $3.91 \times 3.91 \text{ mm}^2$ ; TR/TE: 1.47/0.61 ms; flip angle:  $5^\circ$ ; FOV:  $50 \times 50 \times 36 \text{ cm}^3$ ; receiver bandwidth: 1565 Hz/px). TWIST (Time-resolved angiography With Interleaved Stochastic Trajectories)<sup>44</sup> is a dynamic 3D GRE sequence using view-sharing, where the center of  $k$ -space is sampled more frequently than the periphery in a semi-randomized fashion. Parallel imaging (GRAPPA) and partial Fourier imaging (sampling of 78% in frequency-encoding direction) were used as additional acceleration techniques to further shorten the image acquisition time. To account for geometrical distortions, the manufacturer's correction methods were applied to 3DMRI and 4DMRI. The GTV was contoured on 3DMRI and approved by a trained radiation oncologist.

The first imaging session for each patient was defined as the reference imaging session and taken as a surrogate for the imaging that would be acquired for RT planning. The consecutive imaging sessions  $n \in [2, \dots, N]$ , where  $N$  is the total number of MRI sessions for the patient, were taken as surrogates for the treatment sessions over the course of therapy.

### 2.B. Study workflow

To generate 4DMRI-based target volumes, the GTV position at every point in time of 4DMRI is needed. An overview of the overall workflow performed for this purpose and the subsequent evaluation is given in Fig. 1. The main steps (indicated by the corresponding numbers in Fig. 1) are:

1. Determination of the 3D dataset within 4DMRI that closest resembles the breathing phase of 3DMRI, labeled 4DMRI( $t'$ ),
2. warping of the GTV from 3DMRI to 4DMRI( $t'$ ) by deformable image registration (DIR),
3. propagation of the GTV to all 3D volumes within 4DMRI using DIR,
4. definition of 4DMRI-based ITVs and the midV,
5. creation of PTVs by expansion of the ITVs and midV,
6. creation of a time-averaged 4DMRI,
7. rigid registration (RR) focused on the tumor of time-averaged MR images acquired on day  $n$  to the dataset from the reference MRI session,

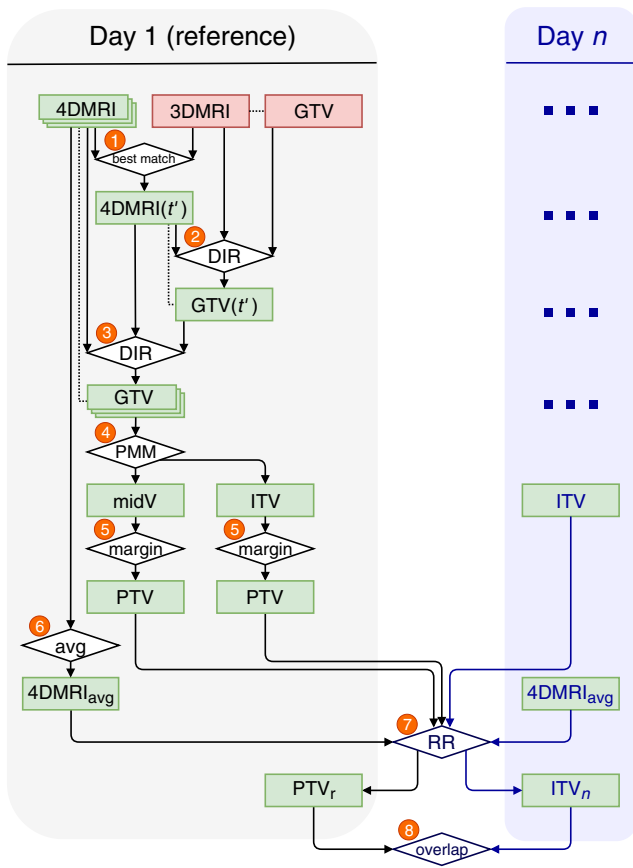


FIG. 1. Overview of data processing and analysis workflow. The individual steps, indicated by the numbers in this figure, are further described in Sections 2.C and 2.D. The dots in the field of day  $n$  indicate that the same workflow steps as shown for day 1 (reference) are performed.  $PTV_r$  represents the reference PTV and  $ITV_n$  the ITV on day  $n$ . GTV: gross tumor volume; DIR: deformable image registration; PMM: passive motion management; ITV: internal target volume; PTV: planning target volume; RR: rigid registration. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

8. geometrical overlap evaluation of PTVs from the reference day ( $PTV_r$ ) and ITVs from day  $n$  ( $ITV_n$ ).

The individual steps are described in more detail in the following sections.

## 2.C. Definition of 4DMRI-based target volumes

### 2.C.1. Determination of breathing states

The superior-inferior (SI) diaphragm positions in each 3D volume of 4DMRI and 3DMRI were determined as surrogates for the breathing state. A region of interest around the transition between the lung and abdominal tissue in the nontumor-bearing hemithorax was manually selected and converted to a binary image by thresholding using Otsu's method.<sup>45</sup> The thresholded image was summed over the anterior-posterior (AP) and right-left (RL) image axes to create a one-dimensional (1D) signal in SI direction. The diaphragm position was defined at the SI position of the steepest gradient of the 1D signal. The point in time  $t'$  was defined as the

time at which the difference of the absolute diaphragm positions of 3DMRI and any of the 3D images of 4DMRI was minimal (step 1 in Fig. 1). The corresponding 3D image at  $t'$  of 4DMRI,  $4DMRI(t')$ , was used for all subsequent registration steps. This step was performed to find the image 4DMRI ( $t'$ ) that is as similar as possible to 3DMRI to reduce uncertainties of the DIR in the following step.

### 2.C.2. Deformable image registration

3DMRI and 4DMRI were resampled to a grid with an isotropic voxel size of  $2 \times 2 \times 2 \text{ mm}^3$  using linear 3D interpolation for all subsequent registration steps. To obtain the GTV on 4DMRI at  $t'$  ( $GTV(t')$ ), 3DMRI was registered to 4DMRI( $t'$ ) (step 2 in Fig. 1) in a multi-level b-spline DIR with mutual information as similarity metric using the software Plastimatch.<sup>46</sup> The DIR was focused on the GTV, expanded by isotropic margins of a few centimeters. Lower uncertainties were expected for the registration of images acquired with the same sequence and therefore a similar contrast ( $4DMRI(t')$  and  $4DMRI(t_i)$ ) compared to a registration of images acquired with two different sequences (3DMRI and  $4DMRI(t_i)$ ). Therefore, to determine the GTV at each point in time  $t_i$  in 4DMRI,  $GTV(t_i)$ , with  $i \in [1, \dots, 157]$ ,  $4DMRI(t')$  was registered to  $4DMRI(t_i)$ , again using a multi-level b-spline DIR focused on the GTV with mutual information as similarity metric (step 3 in Fig. 1). The centroid positions of  $GTV(t_i)$  at all time steps  $t_i$  were measured to evaluate the motion amplitudes of the tumor in each breathing cycle. The amplitudes were calculated relative to the median exhale position of the GTV, since this position is expected to have a lower variance than the inhale position.<sup>4,23</sup>

### 2.C.3. Probability-of-presence ITV

The binary images  $GTV(t_i)$  were summed over all  $t_i$  and divided by the total number of time steps for normalization. The resulting 3D image has the same spatial resolution as the resampled 4DMRI ( $2 \times 2 \times 2 \text{ mm}^3$  voxel size). The voxel values correspond to the percentage of time in which the GTV was present at the voxels' positions over the whole acquisition time. If the acquisition time is long enough so that the respiratory-induced motion during 4DMRI is representative for the motion on the given day, this percentage becomes a POP. This 3D POP distribution was used to determine 4DMRI-based POP ITVs (step 4 in Fig. 1) for cutoffs of 5% and 10% ( $ITV^{5\%}$  and  $ITV^{10\%}$ ). All voxel values equal to or greater than the cutoff were set to 1, while all other voxel values were set to 0.

### 2.C.4. Conventional ITV and midV

To compare  $ITV^{5\%}$  and  $ITV^{10\%}$  with conventional PMM concepts, further target volumes were defined based on the rt-4DMRI dataset (step 4 in Fig. 1). This step was only performed for the reference imaging sessions:

1. An ITV encompassing all GTV positions over the whole acquisition time of 80 s ( $ITV_r^{80s}$ ), which corresponds to a POP cutoff of 0%.
2. To mimic today's standard 4D-CT-based ITV definition workflow, the whole acquisition time was subdivided in eight 10 s periods. For each period, resembling a single 4D-CT scan, an ITV was created by including all GTV position within this period (ie, at 20 consecutive time points) to create eight  $ITV_r^{10s}$ . A similar approach was previously used by Thomas et al.<sup>33</sup>
3. The midV was defined by determining the centroid of the centroids of all 157 GTV positions and then selecting the GTV position at the point in time with the smallest distance of its centroid to this point. This method is equivalent to the approaches previously applied by Ehrbar et al.<sup>47</sup> and Thomas et al.<sup>48</sup>

### 2.C.5. PTV formation

For each patient, the ITVs and the midV of the reference imaging session were expanded by margins to create  $PTV_r$  (step 5 in Fig. 1). An isotropic margin of 5 mm was used for the expansion of the ITVs based on current clinical practice.<sup>27</sup> For the midV, the van Herk formula<sup>43</sup> was used to calculate the margins  $M_d$  in directions  $d$  (with  $d \in [RL,SI,AP]$ ):

$$M_d = \alpha \sqrt{\Sigma_{\text{setup}}^2 + \Sigma_{\text{BL}}^2 + \Sigma_{\text{del}}^2} + \beta \left( \sqrt{\sigma_{\text{setup}}^2 + \sigma_{\text{BL}}^2 + \sigma_{\text{br},d}^2 + \sigma_p^2} - \sigma_p \right), \quad (1)$$

including the systematic and random setup errors for patient positioning with CBCT ( $\Sigma_{\text{setup}} = \sigma_{\text{setup}} = 0.8$  mm<sup>49</sup>), uncertainties due to baseline shifts over the course of treatment ( $\Sigma_{\text{BL}} = 0.99$  mm and  $\sigma_{\text{BL}} = 1.08$  mm<sup>50</sup>), delineation uncertainties ( $\Sigma_{\text{del}} = 1.7$  mm<sup>47</sup>), the standard deviation of the breathing motion of the GTV in direction  $d$  ( $\sigma_{\text{br},d}$ ) and the Gaussian beam penumbra width in lung ( $\sigma_p = 6.4$  mm<sup>16</sup>). The approximation  $\sigma_{\text{br},d} = A_{\text{cen}}^d/3$  was used, where  $A_{\text{cen}}^d$  is the median motion amplitude of the GTV centroid on the reference day in SI, AP and LR direction.<sup>47</sup> The values  $\alpha=2.5$  and  $\beta=1.64$  were chosen to ensure a minimum of 95% of the prescribed dose to the target for 90% of the patients.<sup>43</sup> The margins  $M_d$  were rounded up to integer millimeter values. The ITVs and the midV were expanded by the margins on a  $1 \times 1 \times 1$  mm<sup>3</sup> isovoxel grid and then resampled to the original  $2 \times 2 \times 2$  mm<sup>3</sup> isovoxel grid. As a result of steps 1-5, 12  $PTV_r$  were created for the reference imaging session of each patient:  $PTV_r^{5\%}$ ,  $PTV_r^{10\%}$ ,  $PTV_r^{80s}$ , eight  $PTV_r^{10s}$  and  $PTV_r^{\text{midV}}$ .

### 2.C.6. Rigid registration (RR)

Time-averaged 4DMRI images ( $4DMRI_{\text{avg}}$ ) were calculated by averaging over the 3D volumes of all time steps  $t_i$  (step 6 in Fig. 1). To mimic patient positioning focused on

the moving target during IGRT treatment, which is inevitably blurred on 3D-CBCT images,<sup>3</sup>  $4DMRI_{n,\text{avg}}$  from day  $n$ , was rigidly registered to the reference image  $4DMRI_{r,\text{avg}}$  (step 7 in Fig. 1) also focused on the target region. The RR ITK implementation in Plastimatch with mutual information as similarity metric was used.

## 2.D. Quantification of target volume overlaps

To evaluate the ability of the different PTVs to correctly predict the GTV positions during the course of treatment, a geometrical volume overlap analysis was performed. The reference PTVs of the first MRI sessions ( $PTV_r$ ) were used as surrogates for the target volumes that would be used for RT treatment planning and the ITV of day  $n$  with a POP of 5% ( $ITV_n^{5\%}$ ) as a surrogate for the real target position during treatment on day  $n$ . A cutoff POP of 5% was chosen with the goal to ensure a minimum dose of 95% to the GTV.<sup>43</sup>

### 2.D.1. Geometrical volume overlap analysis

The binary structures  $PTV_r$  and  $ITV_n^{5\%}$  were compared on a voxel-by-voxel basis. A  $PTV_r$  voxel with a value of 1 is a prediction that parts of the GTV will be present during RT delivery at its position with a nonnegligible probability. An  $ITV_n^{5\%}$  voxel with a value of 1 indicates that on day  $n$ , parts of the GTV are present at its position during at least 5% of the time. The total number of voxels that:

1. have a value of 1 in  $PTV_r$  and  $ITV_n^{5\%}$  was labeled *True Positive* (TP),
2. have a value of 1 in  $PTV_r$  and a value of 0 in  $ITV_n^{5\%}$  was labeled *False Positive* (FP) and
3. have a value of 0 in  $PTV_r$  and a value of 1 in  $ITV_n^{5\%}$  was labeled *False Negative* (FN).

This is illustrated in Fig. 2. The geometrical volume overlap of  $PTV_r$  with  $ITV_n^{5\%}$  was quantified in terms of sensitivity and precision, which are defined as following:

Sensitivity (SE; *true positive rate*):

$$SE = \frac{TP}{TP + FN} = \frac{PTV_r \cap ITV_n^{5\%}}{ITV_n^{5\%}} \quad (2)$$

Precision (PRE; *positive predictive value*):

$$PRE = \frac{TP}{TP + FP} = \frac{PTV_r \cap ITV_n^{5\%}}{PTV_r} \quad (3)$$

The Dice similarity coefficient (DSC) can be expressed as a function of SE and PRE:

$$\begin{aligned} DSC &= \frac{2 \cdot TP}{2 \cdot TP + FP + FN} = \frac{2 \cdot (PTV_r \cap ITV_n^{5\%})}{PTV_r + ITV_n^{5\%}} \\ &= 2 \cdot \frac{SE \cdot PRE}{SE + PRE} \end{aligned} \quad (4)$$

In the context of RT, FN can be interpreted as the extent of target miss, FP as the normal tissue damage and SE and



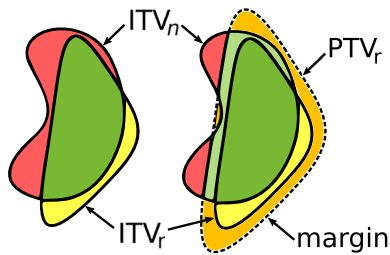


FIG. 2. Illustration of the overlap of  $PTV_r$  and  $ITV_n$ . The assignment of attributes to the different regions is based on the inclusion or exclusion of these regions in  $PTV_r$  and  $ITV_n$ . True positive (TP) areas are marked in dark green, false positive (FP) areas in yellow and false negative (FN) areas in red. In the right image,  $ITV_r$  is expanded by an isotropic margin to obtain  $PTV_r$ , indicated by the dashed line, through which TP is increased and FN decreased (light green area) at the cost of an increased FP (orange area). [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

PRE as the scalar measures of target coverage and normal tissue sparing with respect to the target volume. In an ideal RT treatment, SE and PRE would both be 100%. The general goal of RT is to maximize SE (ie, the tumor coverage) while keeping PRE (ie, the normal tissue sparing) at an acceptable level. The margins around the ITVs and midV to create the PTV increase the SE while PRE decreases. The  $(N-1)$  volume overlaps of each  $PTV_r$  and  $ITV_n^{5\%}$  were evaluated in terms of SE, PRE and DSC (step 8 in Fig. 1). For the  $PTV_r^{10s}$ , the “best”  $PTV_r^{10s}$  (highest SE) and “worst”  $PTV_r^{10s}$  (lowest SE) were determined for each patient.

### 3. RESULTS

#### 3.A. Motion amplitudes

The median number of breathing cycles recorded in the 4DMRI sessions was 15.5 (range: [11.5, 24]) over an acquisition time of 80 s. This corresponds to a median breathing rate of approximately 12 cycles per minute (range: [9, 18]). MRI sessions with stable breathing amplitudes and frequency, as well as sessions with irregular breathing, were observed for all patients. Figure 3(a) shows the GTV centroid motion split up into the SI, RL and AP components for one exemplary MRI session. Figures 3(b)–3(d) depict the motion amplitudes with respect to the median exhale position for all MRI sessions for each patient. The median motion amplitudes are reported in Table I. For all patients and MRI sessions combined, the largest GTV centroid motion was observed in SI direction with a median motion amplitude of 8.8 mm, followed by RL direction (2.7 mm) and AP direction (2.2 mm).

#### 3.B. Probability-of-presence ITV and midV

An exemplary POP distribution is shown in Fig. 4. The gradient of the POP distribution is steep in directions with small motion amplitudes (RL and AP) and shallow in the direction of large motion amplitudes (SI). The median volume ratio of  $ITV^{80s}$  to GTV is reported in Table I. Since this value correlates with the motion amplitude, the largest median value of  $ITV^{80s}/GTV$  was obtained for Patient 2

(3.25), the lowest for Patient 1 (1.29). Averaged over all patients, the mean 3D distance between the centroid of the centroids of all 157 GTV positions and the midV centroid was 0.5 mm.

#### 3.C. PTV formation

All reference ITVs were expanded by isotropic 5 mm margins. For the reference midV, the direction-dependent term  $\sigma_{br,d}$  in the van Herk formula yielded anisotropic margins. The median motion amplitudes during the reference imaging session were 3.5, 0.7, and 2.5 mm in RL direction and 1.8, 0.8, and 3.2 mm in AP direction for Patients 1, 2, and 3, respectively (cf. Fig. 3), which yielded 6 mm margins in both directions. The median SI motion amplitudes of 5.5, 13.2, and 15.7 mm for Patients 1, 2, and 3, respectively, yielded PTV margins of 6, 8, and 9 mm in SI direction.

#### 3.D. Geometrical volume overlap analysis

The results of the geometrical overlap analysis of the different  $PTV_r$  with the  $ITV_n^{5\%}$  are depicted in Fig. 5. Table II summarizes the results and includes the volumes of the  $PTV_r$  relative to the volume of  $PTV_r^{80s}$ .

By definition, the volume of  $PTV_r^{80s}$  was the largest, representing the most conservative motion management approach of the analyzed  $PTV_r$ , where also single extreme GTV positions are included in the ITV. Consequently, the SE values were the largest (median SE between 98–100% for all patients) and the PRE values the lowest of all analyzed  $PTV_r$ .

The variance of the overlap values for the eight  $PTV_r^{10s}$  of the different 10 s periods depends on the regularity of the breathing pattern during the reference imaging session. Since regular breathing patterns were measured for Patients 1 and 3 [cf. Fig. 3(a)], the differences between the overlap values of the best (highest SE) and worst (lowest SE)  $PTV_r^{10s}$  were small ( $\leq 2\%$ ). As the breathing pattern of Patient 2 was irregular during the reference imaging session, the variance of the different  $PTV_r^{10s}$  was larger. A difference of 19% between the SE of the best and worst  $PTV_r^{10s}$  was measured (99% vs 80%). Compared to  $PTV_r^{80s}$ , the same median SE values (differences  $< 1\%$ ) at higher PRE (+4–6%) due to the  $PTV_r$  volume reduction of up to 11% could be achieved for the best  $PTV_r^{10s}$  for all patients. The differences in SE between the  $PTV_r^{80s}$  and the worst  $PTV_r^{10s}$  ranged from 0% (Patient 1) to 19% (Patient 2).

The volumes of  $PTV_r^{midV}$  were the smallest of all  $PTV_r$  with a volume reduction between 16% and 56% compared to  $PTV_r^{80s}$  which led to the highest observed median PRE values ( $\approx 67\%$  for all patients) that were 10–25% higher than for  $PTV_r^{80s}$ . The median SE values for  $PTV_r^{midV}$  were the smallest of all  $PTV_r$  for the three patients, with the largest difference for Patient 2, where a median SE of 68% was measured (compared to 99% for  $PTV_r^{80s}$ ).

For  $PTV_r^{5\%}$ , a median PTV volume reduction with respect to  $PTV_r^{80s}$  of 9–31% was achieved at a similar median SE for Patients 1 and 3 (differences below 2%) but a reduced median



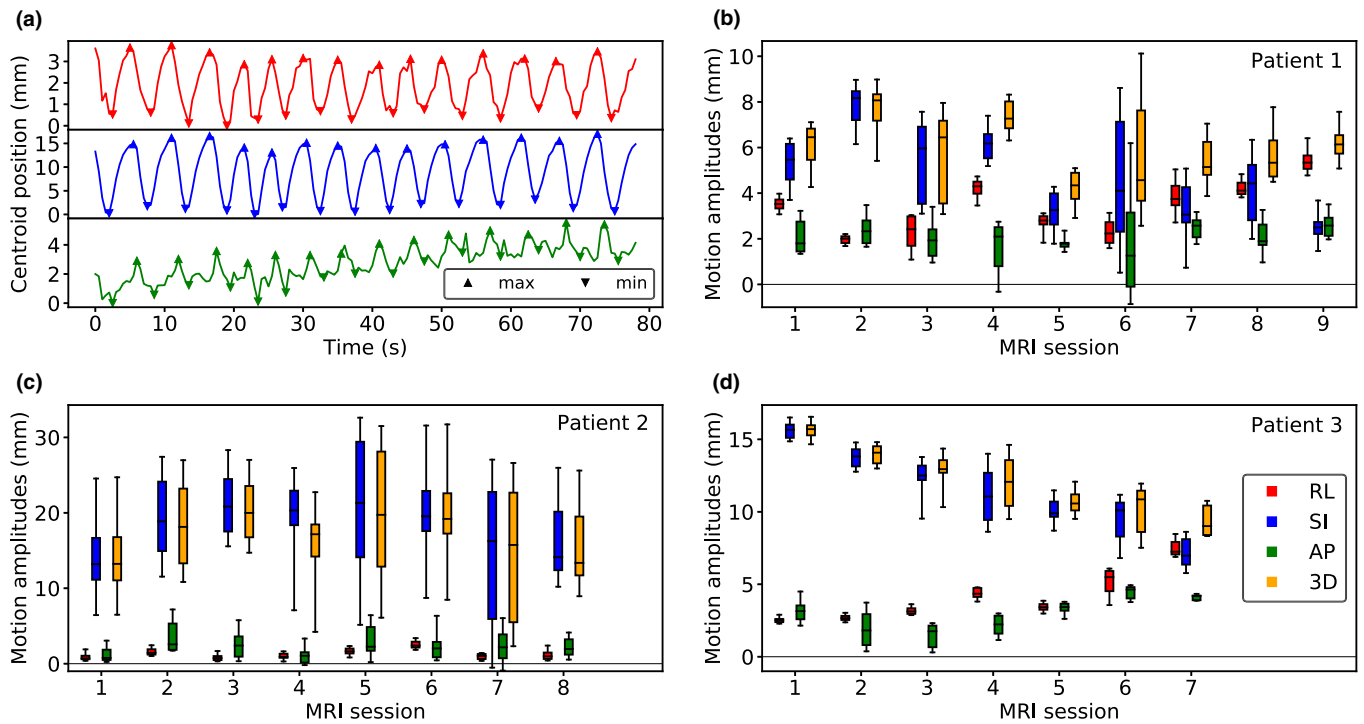


FIG. 3. Gross tumor volume centroid motion amplitudes. (a) Centroid position as a function of time for an exemplary magnetic resonance imaging (MRI) session. The triangular markers indicate the detected maxima and minima for each breathing cycle that were used to calculate the motion amplitudes in right–left (RL) (red), superior–inferior (SI) (blue), and anterior–posterior (AP) (green) directions. (b–d) Motion amplitudes in RL, SI and AP direction and in 3D for all MRI sessions and patients. The whiskers of the boxplots indicate the 5th and 95th percentiles. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE I. Motion parameters for all patients.  $A_{cen}^d$  is the median motion amplitude of the gross tumor volume (GTV) centroid of all four-dimensional magnetic resonance imaging (4DMRI) sessions in three-dimensional (3D), right–left, superior–inferior, and anterior–posterior direction.

Patient	GTV position	$A_{cen}^{3D}$ (mm)	$A_{cen}^{RL}$ (mm)	$A_{cen}^{SI}$ (mm)	$A_{cen}^{AP}$ (mm)	GTV (ml)	$\frac{ITV_{GTV}^{90\%}}{GTV}$
1	Middle right lobe	6.0	3.3	4.4	2.1	273	1.29
2	Lower right lobe	17.2	1.2	18.4	2.0	16	3.25
3	Lower right lobe	12.1	3.4	11.5	3.0	190	1.58

SE for Patient 2 (90% compared to 99%). The SE values for  $PTV_r^{5\%}$  were higher than for  $PTV_r^{midV}$  by up to 22% at the cost of reduced PRE (ie, larger  $PTV_r$  volumes) by up to 11%. The use of  $PTV_r^{10\%}$  instead of  $PTV_r^{5\%}$  showed only small improvements of the PRE ( $\leq 4\%$ ) while the SE dropped by up to 7%.

The behavior of SE and PRE over time is mainly influenced by interfractional changes such as tumor shrinkage and different breathing patterns. The absolute variance of the median motion amplitudes for the different days  $n$  was low for Patient 1 (standard deviation of median 3D centroid motion amplitude was 1.3 mm) and the median motion amplitudes were decreasing over time for Patient 3. In combination with the regression or stagnation of the GTV size over time that was observed for these two patients, this led to a stable SE for all days  $n$  and all  $PTV_r$  (standard deviation of SE values  $< 4\%$  for both patients and all  $PTV_r$ ).

In contrast, the interfractional variance of the SE of the different  $PTV_r$  for Patient 2 was considerably larger (up to

10% difference for  $PTV_r^{midV}$ ; cf. Fig. 6). This is a consequence of the high absolute variance of the median motion amplitudes for different days  $n$  [cf. Fig. 3(c)]. While the median GTV centroid motion amplitudes in the first and last MR imaging session were equivalent (13 mm), the amplitudes in the remaining imaging sessions were substantially larger (16–20 mm). This led to markedly higher SE values for the last compared to the other imaging sessions for most of the  $PTV_r$  (cf. Fig. 6). The PRE for Patient 2 was gradually decreasing over time for all  $PTV_r$ , which can be explained by the tumor shrinkage that was observed for this patient.

#### 4. DISCUSSION

The largest motion amplitudes were observed in SI direction. Considerably larger motion amplitudes were measured for GTVs in the lower lobe (Patients 2 and 3) than in the middle lobe (Patient 1). These findings are consistent with observations described in literature.<sup>4,9,51,52</sup>

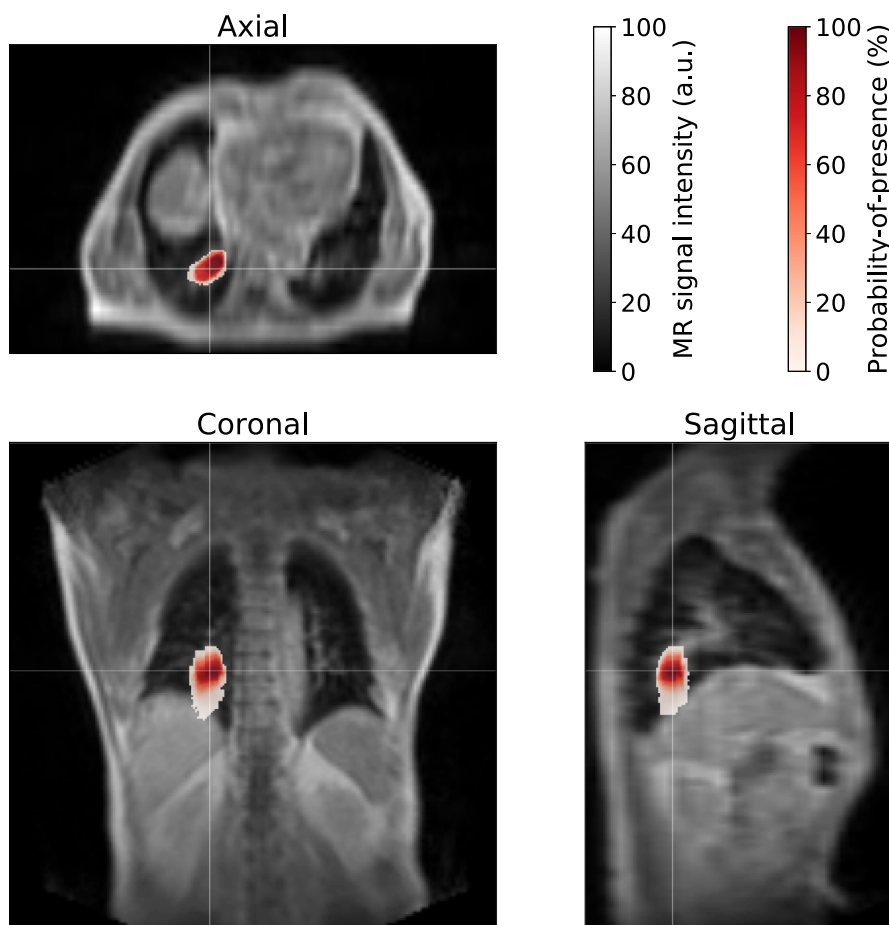


FIG. 4. Three-dimensional probability-of-presence (POP) distribution. The POP is color-encoded and superimposed on four-dimensional-magnetic resonance imaging (MRI) ( $t'$ ) for the reference MRI session of Patient 2. The crosshairs indicate the slice positions in the respective axial, coronal, and sagittal views. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

The  $PTV_r^{80s}$ , as the most conservative PMM approach investigated in this study, yielded the highest SE values at the largest PTV volumes that might be unacceptably large for clinical application.

The large difference of the SE between the best and worst  $PTV_r^{10s}$  that was observed for Patient 2 demonstrates the risk of sampling the breathing motion over only a short time period. When an ITV is defined based on this motion sample—as is done routinely today in the 4D-CT-based ITV workflow—intrafractional uncertainties would directly translate to systematic errors impacting the whole RT treatment.

The SE values for Patient 1 were high for all  $PTV_r$  and all imaging sessions. This patient would likely benefit from being treated with the mid-ventilation approach, as the  $PTV_r^{midV}$  was smaller and hence the PRE was larger than for the ITV-based  $PTV_r$ . The low SE values of most  $PTV_r$  for Patient 2 indicate that the respiratory-induced GTV motion captured on the reference day was not representative for the motion of the GTV in the remaining imaging sessions. In particular, the SE of  $PTV_r^{midV}$  and  $ITV_n^{5\%}$  was as low as 68%. A plausible explanation for this result can be inferred from Fig. 3(c). The variance of the SI GTV centroid motion amplitudes during the reference imaging session was high. The

breathing motion was considered as a random uncertainty in the calculation of the PTV margin for the midV. However, since only the median motion amplitude was used for the margin determination, the rich information about the complex GTV motion trajectory was lost in this simplification step. Although large SI amplitudes of about twice the median value were observed during the reference imaging session, this was not appropriately accounted for by the PTV margins which in turn led to the poor SE. The SE values for Patient 3 were higher than 95% for all  $PTV_r$  with the exception of  $PTV_r^{midV}$  for which a SE of 90% was measured at a volume reduction of 23% relative to  $PTV_r^{80s}$ . Without a dosimetric analysis it cannot be concluded whether a SE value of 90% would be clinically acceptable.

The overall variance of the SE for all patients accumulated was second lowest for  $PTV_r^{5\%}$  (after  $PTV_r^{80s}$ ) at a high median SE value, indicating a higher robustness of this approach compared to  $PTV_r^{10s}$  and  $PTV_r^{midV}$  (cf. Fig. 5). While the SE for  $PTV_r^{5\%}$  was reduced for Patient 2, it was not as low as for  $PTV_r^{midV}$ . By using the full information of the GTV motion over the whole duration of the reference imaging session, random uncertainties in the PTV definition could be reduced through the use of  $ITV_r^{5\%}$ .

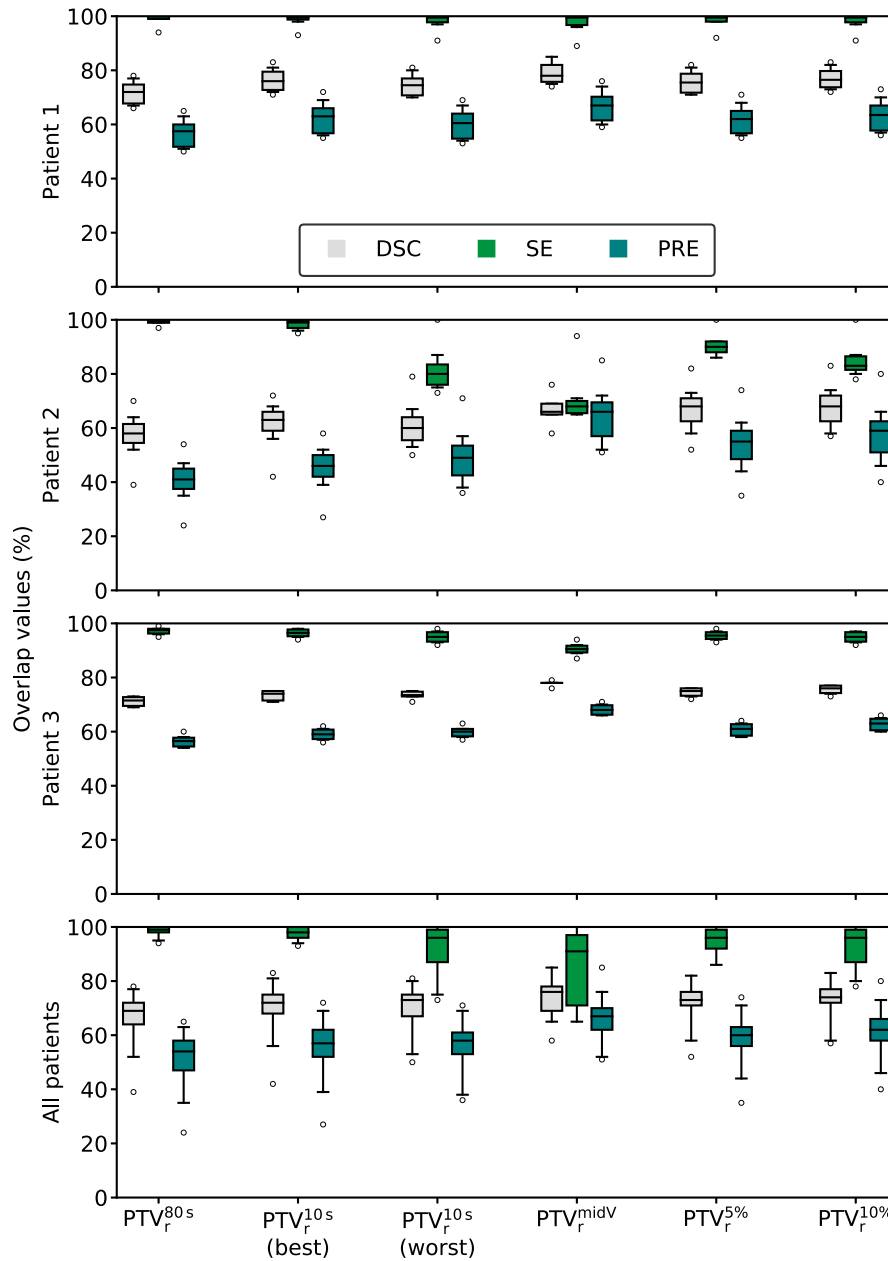


FIG. 5. Geometrical overlap analysis for different passive motion-management concepts. The overlap values are plotted for each of the three patients separately and for all patients accumulated. Each boxplot contains the values for all follow-up imaging sessions.  $PTV_r^{10s}$  (best/worst) are defined as the  $PTV_r^{10s}$  with the highest/lowest SE value. The whiskers of the boxplots indicate the 5th and 95th percentiles. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE II. Summary of geometrical overlap analysis. The median overlap values over all follow-up imaging sessions are given in percent for the different  $PTV_r$  and patients.  $V_{rel}$  describes the volume of the different  $PTV_r$  relative to the volume of  $PTV_r^{80s}$  in percent.

	Patient 1				Patient 2				Patient 3			
	DSC	SE	PRE	$V_{rel}$	DSC	SE	PRE	$V_{rel}$	DSC	SE	PRE	$V_{rel}$
$PTV_r^{80s}$	72	100	57	100	58	99	41	100	72	98	56	100
$PTV_r^{10s}$ (best)	76	100	63	90	63	99	46	89	75	97	60	93
$PTV_r^{10s}$ (worst)	75	100	61	94	60	80	49	67	74	95	60	92
$PTV_r^{midV}$	78	99	67	84	66	68	66	44	78	90	68	77
$PTV_r^{5\%}$	75	100	62	91	68	90	55	69	75	96	61	91
$PTV_r^{10\%}$	77	100	64	88	68	83	59	60	76	95	63	87

DSC, Dice similarity coefficient; SE, sensitivity; PRE, precision.

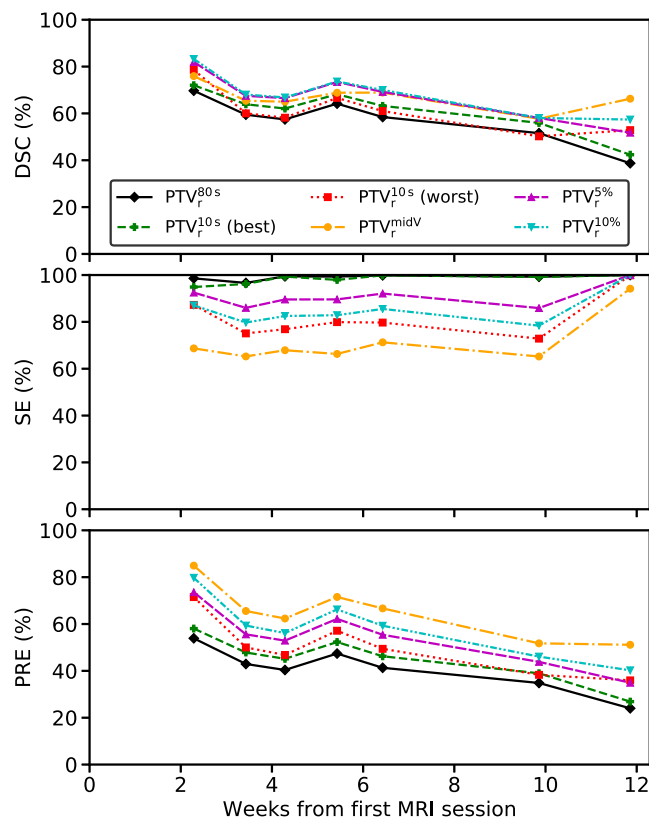


FIG. 6. Geometrical overlap values over time. For Patient 2, the overlap values for the six different  $PTV_r$  with  $ITV_n^{5\%}$  are plotted for all imaging sessions on days  $n$ . [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

Studies by Ehrbar et al.<sup>47</sup> and Thomas et al.<sup>48</sup> concluded in their dosimetric analyses that the mid-ventilation approach can reduce the dose to the lung with no or only slightly reduced tumor coverage with respect to the ITV approach. Both studies however, recalculated the dose on the same 4D-CT images that were used for definition of the target volumes. The results of the present study, however, suggest that the SE could be strongly affected by interfractional changes and the mid-ventilation concept could be less robust than the more conservative  $ITV_r^{5\%}$  concept in these cases.

Keeping the volumetric effect in mind, setting a reasonable POP cutoff can lead to an over-proportionate improvement of PRE while the SE decreases less. However, setting this cutoff higher bears the risk of creating underdosed (“cold”) spots at the edges of the GTV. Finding an ideal POP cutoff without an extensive dosimetric evaluation is not a trivial task and the optimum is expected to be patient-specific. For the patients included in this study, the use of a POP cutoff of 10% ( $PTV_r^{10\%}$ ) instead of 5% resulted in only a slight improvement of PRE at a reduced SE. As a high SE (surrogate for target coverage) is of higher importance than a high PRE (surrogate for normal tissue sparing), the use of  $PTV_r^{10\%}$  instead of  $PTV_r^{5\%}$  to improve the PRE would not be justified in these cases.

For all patients, the differences of SE and PRE between the analyzed  $PTV_r$  were larger than for the DSC. Due to the different importances of SE and PRE, the interpretation of

the overlap results based on the DSC alone was found not to be sufficient to evaluate which  $PTV_r$  represented the optimal target volume.

Different gradients of the POP distribution in different directions were observed depending on the motion amplitudes in the respective directions. This information could be used to create anisotropic safety margins around  $ITV_r^{5\%}$  in a similar way as done for the margin expansion of the midV. Under the assumption that adequate image guidance techniques are used, this patient-specific margin could potentially increase the PRE while keeping the SE and hence the risk of underdosage of the GTV at a constant level.

A potential extension of the concepts presented in this study could be to directly incorporate the POP information in the treatment planning process. Following the work published by Shusharina et al.,<sup>53</sup> an “internal target distribution” based on the voxel-individual POP values could be defined instead of a binary ITV as input for a probabilistic treatment planning approach.

The results for Patient 2 indicate that most  $PTV_r$  were not representative for the GTV motion in the following imaging sessions, resulting in a reduced SE for sessions 2–7. With regard to the use of SE and PRE in adaptive RT,<sup>5,6</sup> if the overlap parameters showed a clear deviation or decrease over time, this would indicate that the patient would benefit from replanning. These findings are supported by the conclusions drawn by St. James et al.<sup>19,24</sup> who demonstrated that a low SE (called “ITV coverage” in their study) could potentially lead to substantial underdosage of the target.

The presented method could be adapted for the use of retrospectively sorted rc-4DMRI instead of rt-4DMRI. rc-4DMRI provides higher spatial resolution than rt-4DMRI, but since only a single breathing cycle is reconstructed in rc-4DMRI, weighting factors for the probability of occurrence of the different breathing phases would have to be determined to reconstruct a 3D POP distribution. Uncertainties associated with the retrospective sorting would have to be accounted for. When 2D images are acquired in rc-4DMRI, instead of 3D images as in rt-4DMRI, less motion information per time is sampled. To achieve the same statistical uncertainty level compared to rt-4DMRI, longer acquisition times would be necessary.

A future workflow for the integration of the proposed 4DMRI-based POP ITV concept into clinical practice could be based on the following steps:

1. Acquisition of a 4DMRI scan instead or in addition to the 4D-CT scan at the planning stage,
2. definition of a 4DMRI-based ITV based on the POP of the GTV by choosing a suitable cutoff probability to overcome current limitations of a 4D-CT based ITV,
3. definition of additional margins around the ITV to prospectively account for potential interfractional changes,
4. use of image-guidance techniques for patient positioning (focused on the target region) and to detect relevant interfractional changes during RT treatment,

5. regular repetition of the 4DMRI scan and subsequent target volume overlap analysis with the parameters SE and PRE over the course of the RT treatment to assess interfractional changes in breathing motion and to verify the applied margin concept,
6. decision whether the original plan is still appropriate or replanning is necessary based on the overlap analysis, ideally in conjunction with a dosimetric evaluation.

This workflow would not necessarily replace the current 4D-CT-based workflow, but could be an enhancement for increased treatment accuracy and quality assurance, especially for patients who experience substantial interfractional changes.

In this proof-of-concept study, the patient cohort was limited to three patients, but compared to other studies that have evaluated lung tumor motion with MRI (cf. Table I in the publication by Thomas et al.<sup>33</sup>), the number of MRI sessions per patient is higher by a factor of 2-4. The acquisition time of the 4DMRI was limited to 80 s due to technical reasons, which is shorter than 2D cine-MR imaging studies reported in literature.<sup>33-35</sup> However, the proposed method could be easily extended to longer acquisition times in the future to improve statistics to assure that the representative target motion of the day is captured.<sup>22</sup> The DIR steps introduce uncertainties that are difficult to quantify but that are expected to be counterbalanced by the higher representativity of the target motion description compared to a 4D-CT-based ITV definition. Potential geometrical image distortions of the MR images have to be accounted for. The proposed workflow is more labor-intensive than today's clinical routine workflow but has the potential to reduce the integral dose to the patient and enhance target coverage. A dosimetric evaluation of the effect of the changing overlap values of the ITVs of different days was beyond the scope of this study. The spatial and temporal resolution of today's rt-4DMRI sequences is limited, including the sequence used in this study. A limited spatial resolution affects the accuracy of the image registration and the target delineation. To account for this, related errors could be estimated and absorbed in the PTV margin calculation.<sup>47</sup> Cai et al.<sup>22</sup> showed that a frame rate of fewer than two images per second could affect the reproducibility of the POP. A limited temporal resolution can furthermore lead to an apparent enlargement of moving structures and an underestimation of the inhale and exhale positions.<sup>42</sup> The former effect would lead to an overly conservative estimation of the GTV POP and therefore decrease the PRE. The latter effect could potentially lead to an underdosage of the GTV edges (lower SE). Since on average ten 3D images were acquired per breathing cycle, the underestimation of the range of motion is estimated to be similar or less pronounced compared to a 4D-CT scan that consists of ten phases. However, this effect has to be considered for patients with high breathing frequencies (>12 cycles/min). Improvements of the spatial and temporal resolution of rt-4DMRI are expected in the near future.<sup>39</sup>

## 5. CONCLUSIONS

We proposed and investigated a novel concept for ITV definition based on rt-4DMRI and pointed out its potentials and limitations. This ITV definition is based on the POP of the target in 3D, which is expanded by a PTV margin to account for interfractional anatomical and motion changes. In combination with image guidance techniques, the proposed method has the potential to reduce the statistical and systematic uncertainties associated with today's clinical standard-of-care workflow based on ITVs or midV defined on 4D-CT scans. Hereby, an improved target coverage, balanced against the dose to adjacent normal tissues, could be achieved. While this study was focused on the motion of lung tumors, the methods could be translated to other tumor sites which are strongly affected by intrafractional and interfractional motion, such as the liver or pancreas.

## ACKNOWLEDGMENTS

This work was supported by the German Research Foundation (DFG) within the Research Training Group GRK 2274.

## CONFLICT OF INTEREST

The authors have no conflict to disclose.

\*Both authors share last authorship.

<sup>†</sup>Author to whom correspondence should be addressed. Electronic mail: moritz.rabe@med.uni-muenchen.de.

## REFERENCES

1. Li H, Chang JY. Accounting for, mitigating, and choice of margins for moving tumors. *Semin Radiat Oncol.* 2018;28:194-200.
2. Knybel L, Cvek J, Molenda L, Stieberova N, Feltl D. Analysis of lung tumor motion in a large sample: patterns and factors influencing precise delineation of internal target volume. *Int J Radiat Oncol Biol Phys.* 2016;96:751-758.
3. Cailliet V, Booth JT, Keall P. IGRT and motion management during lung SBRT delivery. *Phys Med.* 2017;44:113-122.
4. Brandner ED, Chetty IJ, Giaddui TG, Xiao Y, Huq MS. Motion management strategies and technical issues associated with stereotactic body radiotherapy of thoracic and upper abdominal tumors: a review from NRG oncology. *Med Phys.* 2017;44:2595-2612.
5. Menten MJ, Wetscherek A, Fast MF. MRI-guided lung SBRT: present and future developments. *Phys Med.* 2017;44:139-149.
6. Paganelli C, Whelan B, Peroni M, et al. MRI-guidance for motion management in external beam radiotherapy: current status and future challenges. *Phys Med Biol.* 2018;63:22TR03.
7. Yang M, Timmerman R. Stereotactic ablative radiotherapy uncertainties: delineation, setup and motion. *Semin Radiat Oncol.* 2018;28:207-217.
8. Dhont J, Vandemeulebroucke J, Burghelma M, 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;126:339-346.
9. Keall PJ, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Med Phys.* 2006;33:3874-3900.
10. De Ruysscher D, Faivre-Finn C, Moeller D, et al. European Organization for Research and Treatment of Cancer (EORTC) recommendations for planning and delivery of high-dose, high precision radiotherapy for lung cancer. *Radiother Oncol.* 2017;124:1-10.



11. Grégoire V, Mackie T, Neve W, et al. ICRU Report 83: prescribing, recording, and reporting photon-beam intensity-modulated radiation therapy (IMRT). *J ICRU*. 2010;10:1–112.
12. Pan T, Lee TY, Rietzel E, Chen GTY. 4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT. *Med Phys*. 2004;31:333–340.
13. Tseng YD, Wootton L, Nyflot M, et al. 4D computed tomography scans for conformal thoracic treatment planning: is a single scan sufficient to capture thoracic tumor motion? *Phys Med Biol*. 2018;63:02NT03.
14. Wolthaus JWH, Sonke JJ, van Herk M, Damen EMF. Reconstruction of a time-averaged midposition CT scan for radiotherapy planning of lung cancer patients using deformable registration. *Med Phys*. 2008;35:3998–4011.
15. Wolthaus JW, Schneider C, Sonke JJ, et al. Mid-ventilation CT scan construction from four-dimensional respiration-correlated CT scans for radiotherapy planning of lung cancer patients. *Int J Radiat Oncol Biol Phys*. 2006;65:1560–1571.
16. Wolthaus JW, Sonke JJ, van Herk M, et al. Comparison of different strategies to use four-dimensional computed tomography in treatment planning for lung cancer patients. *Int J Radiat Oncol Biol Phys*. 2008;70:1229–1238.
17. Mercieca S, Belderbos JS, Jaeger KD, et al. Interobserver variability in the delineation of the primary lung cancer and lymph nodes on different four-dimensional computed tomography reconstructions. *Radiother Oncol*. 2018;126:325–332.
18. Guckenberger M, Wilbert J, Meyer J, Baier K, Richter A, Flentje M. Is a single respiratory correlated 4D-CT study sufficient for evaluation of breathing motion? *Int J Radiat Oncol Biol Phys*. 2007;67:1352–1359.
19. St. James S, Mishra P, Hacker F, Berbeco RI, Lewis JH. Quantifying ITV instabilities arising from 4DCT: a simulation study using patient data. *Phys Med Biol*. 2012;57:L1–L7.
20. Harada K, Katoh N, Suzuki R, et al. Evaluation of the motion of lung tumors during stereotactic body radiation therapy (SBRT) with four-dimensional computed tomography (4DCT) using real-time tumor-tracking radiotherapy system (TRTR). *Phys Med*. 2016;32:305–311.
21. Steiner E, Shieh CC, Caillet V, et al. Both four-dimensional computed tomography and four-dimensional cone beam computed tomography under-predict lung target motion during radiotherapy. *Radiother Oncol*. 2019;135:65–73.
22. Cai J, Read PW, Larner JM, Jones DR, Benedict SH, Sheng K. Reproducibility of interfraction lung motion probability distribution function using dynamic MRI: statistical analysis. *Int J Radiat Oncol Biol Phys*. 2008;72:1228–1235.
23. Seppenwoolde Y, Shirato H, Kitamura K, et al. Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat. *Int J Radiat Oncol Biol Phys*. 2002;53:822–834.
24. St. James S, Seco J, Mishra P, Lewis JH. Simulations using patient data to evaluate systematic errors that may occur in 4D treatment planning: a proof of concept study. *Med Phys*. 2013;40:091706.
25. Green OL, Rankine LJ, Cai B, et al. First clinical implementation of real-time, real anatomy tracking and radiation beam control. *Med Phys*. 2018;45:3728–3740.
26. van Herk M, McWilliam A, Dubec M, Faivre-Finn C, Choudhury A. Magnetic resonance imaging-guided radiation therapy: a short strengths, weaknesses, opportunities, and threats analysis. *Int J Radiat Oncol Biol Phys*. 2018;101:1057–1060.
27. Nabavizadeh N, Elliott DA, Chen Y, et al. Image guided radiation therapy (IGRT) practice patterns and IGRT's impact on workflow and treatment planning: results from a National Survey of American Society for Radiation Oncology members. *Int J Radiat Oncol Biol Phys*. 2016;94:850–857.
28. Sweeney RA, Seubert B, Stark S, et al. Accuracy and inter-observer variability of 3D versus 4D cone-beam CT based image-guidance in SBRT for lung tumors. *Radiat Oncol*. 2012;7:81.
29. Brock KK, Dawson LA. Point: principles of magnetic resonance imaging integration in a computed tomography-based radiotherapy workflow. *Semin Radiat Oncol*. 2014;24:169–174.
30. Koch N, Liu H, Starkschall G, et al. Evaluation of internal lung motion for respiratory-gated radiotherapy using MRI: part I – correlating internal lung motion with skin fiducial motion. *Int J Radiat Oncol Biol Phys*. 2004;60:1459–1472.
31. Blackall JM, Ahmad S, Miquel ME, McClelland JR, Landau DB, Hawkes DJ. MRI-based measurements of respiratory motion variability and assessment of imaging strategies for radiotherapy planning. *Phys Med Biol*. 2006;51:4147–4169.
32. Tryggestad E, Flammang A, Hales R, et al. 4D tumor centroid tracking using orthogonal 2D dynamic MRI: implications for radiotherapy planning. *Med Phys*. 2013;40:091712.
33. Thomas DH, Santhanam A, Kishan AU, et al. Initial clinical observations of intra- and interfractional motion variation in MR-guided lung SBRT. *Br J Radiol*. 2018;91:20170522.
34. Cusumano D, Dhont J, Boldrini L, et al. Predicting tumour motion during the whole radiotherapy treatment: a systematic approach for thoracic and abdominal lesions based on real time MR. *Radiother Oncol*. 2018;129:456–462.
35. van Sörnsen de Koste JR, Palacios MA, Bruynzeel AM, Slotman BJ, Senan S, Lagerwaard FJ. MR-guided gated stereotactic radiation therapy delivery for lung, adrenal, and pancreatic tumors: a geometric analysis. *Int J Radiat Oncol Biol Phys*. 2018;102:858–866.
36. Cai J, Read PW, Baisden JM, Larner JM, Benedict SH, Sheng K. Estimation of error in maximal intensity projection-based internal target volume of lung tumors: a simulation and comparison study using dynamic magnetic resonance imaging. *Int J Radiat Oncol Biol Phys*. 2007;69:895–902.
37. Paganelli C, Lee D, Kipritidis J, et al. Feasibility study on 3D image reconstruction from 2D orthogonal cine-MRI for MRI-guided radiotherapy. *J Med Imaging Radiat Oncol*. 2018;62:389–400.
38. von Siebenthal M, Székely G, Gamper U, Boesiger P, Lomax A, Cattin P. 4D MR imaging of respiratory organ motion and its variability. *Phys Med Biol*. 2007;52:1547–1564.
39. Stemkens B, Paulson ES, Tjijssen RHN. Nuts and bolts of 4D-MRI for radiotherapy. *Phys Med Biol*. 2018;63:21TR01.
40. Kauczor H-U, Plathow C. Imaging tumour motion for radiotherapy planning using MRI. *Cancer Imaging*. 2006;6:S140–S144.
41. Dinkel J, Hintze C, Tetzlaff R, et al. 4D-MRI analysis of lung tumor motion in patients with hemidiaphragmatic paralysis. *Radiother Oncol*. 2009;91:449–454.
42. Biederer J, Hintze C, Fabel M, Dinkel J. Magnetic resonance imaging and computed tomography of respiratory mechanics. *J Magn Reson Imaging*. 2010;32:1388–1397.
43. van Herk M, Remeijer P, Rasch C, Lebesque JV. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. *Int J Radiat Oncol Biol Phys*. 2000;47:1121–1135.
44. Laub G, Kroeker R. syngo TWIST for dynamic time-resolved MR angiography. *Magnetom Flash*. 2006;3:92–95.
45. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern*. 1979;9:62–66.
46. Shackelford JA, Kandasamy N, Sharp GC. On developing B-spline registration algorithms for multi-core processors. *Phys Med Biol*. 2010;55:6329–6351.
47. Ehrbar S, Jöhl A, Tartas A, et al. ITV, mid-ventilation, gating or couch tracking – a comparison of respiratory motion-management techniques based on 4D dose calculations. *Radiother Oncol*. 2017;124:80–88.
48. Thomas SJ, Evans BJ, Harihar L, Chantler HJ, Martin AG, Harden SV. An evaluation of the mid-ventilation method for the planning of stereotactic lung plans. *Radiother Oncol*. 2019;137:110–116.
49. Garibaldi C, Piperno G, Ferrari A, et al. Translational and rotational localization errors in cone-beam CT based image-guided lung stereotactic radiotherapy. *Phys Med*. 2016;32:859–865.
50. Lang S, Shrestha B, Graydon S, et al. Clinical application of flattening filter free beams for extracranial stereotactic radiotherapy. *Radiother Oncol*. 2013;106:255–259.
51. Yu ZH, Lin SH, Balter P, Zhang L, Dong L. A comparison of tumor motion characteristics between early stage and locally advanced stage lung cancers. *Radiother Oncol*. 2012;104:33–38.
52. Schwarz M, Cattaneo GM, Marrazzo L. Geometrical and dosimetric uncertainties in hypofractionated radiotherapy of the lung: a review. *Phys Med*. 2017;36:126–139.
53. Shusharina N, Craft D, Chen Y-L, Shih H, Bortfeld T. The clinical target distribution: a probabilistic alternative to the clinical target volume. *Phys Med Biol*. 2018;63:155001.