

University of Groningen

## Evaluation of interplay and organ motion effects by means of 4D dose reconstruction and accumulation

Meijers, Arturs; Knopf, Antje-Christin; Crijns, Anne P G; Ubbels, Jan F; Niezink, Anne G H; Langendijk, Johannes A; Wijsman, Robin; Both, Stefan

*Published in:*  
Radiotherapy and Oncology

*DOI:*  
[10.1016/j.radonc.2020.07.055](https://doi.org/10.1016/j.radonc.2020.07.055)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2020

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Meijers, A., Knopf, A.-C., Crijns, A. P. G., Ubbels, J. F., Niezink, A. G. H., Langendijk, J. A., Wijsman, R., & Both, S. (2020). Evaluation of interplay and organ motion effects by means of 4D dose reconstruction and accumulation. *Radiotherapy and Oncology*, 150, 268-274. <https://doi.org/10.1016/j.radonc.2020.07.055>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*



## Original Article

## Evaluation of interplay and organ motion effects by means of 4D dose reconstruction and accumulation



Arturs Meijers\*, Antje-Christin Knopf, Anne P.G. Crijns, Jan F. Ubbels, Anne G.H. Niezink, Johannes A. Langendijk, Robin Wijsman, Stefan Both

University of Groningen, University Medical Centre Groningen, Department of Radiation Oncology, Groningen, The Netherlands

## ARTICLE INFO

## Article history:

Received 24 March 2020

Received in revised form 14 July 2020

Accepted 23 July 2020

Available online 5 August 2020

## Keywords:

Pencil beam scanning

Proton therapy

Thoracic cancer

Interplay

4D dose reconstruction

## ABSTRACT

**Purpose:** Pencil beam scanned proton therapy (PBS-PT) treatment quality might be compromised by interplay and motion effects. Via fraction-wise reconstruction of 4D dose distributions and dose accumulation, we assess the clinical relevance of motion related target dose degradation in thoracic cancer patients.

**Methods and materials:** For the ten thoracic patients (Hodgkin lymphoma and non-small cell lung cancer) treated at our proton therapy facility, daily breathing pattern records, treatment delivery log-files and weekly repeated 4DCTs were collected. Patients exhibited point-max target motion of up to 20 mm. They received robustly optimized treatment plans, delivered with five-times rescanning in fractionated regimen. Treatment delivery records were used to reconstruct 4D dose distributions and the accumulated treatment course dose per patient. Fraction-wise target dose degradations were analyzed and the accumulated treatment course dose, representing an estimation of the delivered dose, was compared with the prescribed dose.

**Results:** No clinically relevant loss of target dose homogeneity was found in the fraction-wise reconstructed 4D dose distributions. Overall, in 97% of all reconstructed fraction doses, D98 remained within 5% from the prescription dose. The V95 of accumulated treatment course doses was higher than 99.7% for all ten patients.

**Conclusions:** 4D dose reconstruction and accumulation enables the clinical estimation of actual exhibited interplay and motion effects. In the patients considered here, the loss of homogeneity caused by interplay and organ motion did not show systematic pattern and smeared out throughout the course of fractionated PBS-PT treatment. Dose degradation due to anatomical changes showed to be more severe and triggered treatment adaptations for five patients.

© 2020 The Authors. Published by Elsevier B.V. Radiotherapy and Oncology 150 (2020) 268–274 This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Pencil beam scanned proton therapy (PBS PT) treatments of intrathoracic targets are associated with significant uncertainty. Treatment quality may be compromised by setup errors, range uncertainties, respiratory motion baseline shifts, anatomical changes, delivery inaccuracies and motion of various sources (e.g. respiration, cardiac motion, swallowing, etc.). More specifically, the relative motion between a thoracic target volume and the scanning proton beam can cause deviation of the delivered dose from the planned dose and is referred to as the interplay effect. The interplay effect has been of concern since the clinical introduction of PBS PT [1] and has hampered its wide range clinical deployment for the treatment of thoracic indications [2].

To generate PBS PT plans that are robust against possible uncertainties, robust optimization techniques have been introduced and

are increasingly used in clinical routine, especially when treating moving targets. Outcomes of the robust optimization must be evaluated to check that robustness objectives are met. Robustness evaluation is commonly performed through simulations of multiple error scenarios [3]. The outcome of robustness evaluation depends on the sampling of uncertainties. The more realistic and comprehensive a robustness evaluation is, the more time consuming and computationally expensive it gets, eventually making it unfeasible for deployment in clinical routine on regular basis.

Especially the assessment of the clinical impact of the interplay effect is difficult due to the large parameter space affecting it and has been subject of many in-silico simulation studies [4–8]. Three sets of parameters determine the interplay effect: (i) delivery characteristics in the sense of the delivery timeline, the start of delivery with respect to the organ motion (starting phase), the scanning path, the spot size, the number of applied rescans, the type of rescanning (e.g. layered or volumetric rescanning) and the use of, e.g.,

\* Corresponding author at: Hanzeplein 1, 9713 GZ Groningen, The Netherlands.  
E-mail address: [a.meijers@umcg.nl](mailto:a.meijers@umcg.nl) (A. Meijers).

gating, (ii) plan characteristics such as the field directions, the number of fields and layer spacing and (iii) patient characteristics such as the patient and fraction specific motion pattern in terms of amplitude, frequency and variability and the target volume size and location. Due to this large amount of variables, the impact of the interplay effect for a specific patient so far has only been assessed in a probabilistic manner by simulating several possible delivery scenarios with varying input parameters (e.g. the starting phase, the number of rescans, etc.). In this way, possible deviations of the delivered dose from the planned dose can be estimated. However, it remains unknown what is the probability of one scenario over another and there is lack of insight on what dose actually is delivered to the patient. Consequently, when correlating outputs to the planned treatment dose, this correlation is impeded by the unknown deviation of the actual delivered dose from the planned dose.

Accounting for dose delivery uncertainties by robust optimization, considering many of possible treatment scenarios, comes at the cost of integral dose and dose to healthy tissues. However, for every individual patient only one scenario occurs. By creating overly robust plans, the normal tissue may be unnecessarily overdosed to compensate for situations that may never happen in practice.

Contrasting the concept of “overly conservative” robust optimization is the concept of adaptive treatment delivery [9]. Here, much more conformal treatment plans can be delivered on the basis that deviations from the nominal scenario are accounted for by a plan adaptation. Adaptation can either be triggered based on continuous accumulation and evaluation of the delivered dose distribution or executed on a daily basis or even real-time in the future. However, that requires high degree and reliability of automation. A triggered adaptation relies on an assessment of the actual delivered fraction dose, which in general is challenging to obtain, especially for thoracic indications due to motion affecting the treatment delivery.

We have developed a methodology aimed at gaining more insight on fraction-wise delivered dose distributions. This methodology is based on retrospective reconstruction of four-dimensional (4D) dose distributions and accumulation providing means to continuously assess treatment course quality [10]. 4D dose reconstruction for carbon ion has previously been reported on by Richter et al. [11]. In the current study commercially available solutions have been used to implement the dose reconstruction workflow, different indications and much longer fractionation schemes with multiple repeat CTs have been investigated. Furthermore, investigated treatment plans have been prepared using robust optimization and evaluation planning techniques.

Here we present initial results of the application of 4D dose reconstruction and accumulation (4DREAL) of 10 consecutive patients with thoracic indications (Hodgkin lymphoma (post-chemotherapy) and non-small cell lung cancer (NSCLC)) treated with PBS PT at our facility. The focus of this study is to primarily evaluate fraction-wise and consecutive accumulated target volume doses (high dose area). The objective is to assess the impact of the interplay and organ motion on the target dose homogeneity and to investigate the consequences of fraction-wise loss of homogeneity on the total accumulated course dose.

## Material and methods

In our facility the treatment of targets, which are affected by respiratory motion, is performed following a procedure based on four principles: (1) motion assessment, (2) robust planning, (3) robustness evaluation and (4) retrospective 4D dose reconstruction.

### (1) Motion assessment

Planning 4DCT (phase-based reconstruction) is used to assess the magnitude of the motion. Phase-based reconstructions are more suitable for dose calculations due to equidistant spacing in time, however, may suffer from anatomy-induced artifacts, which could be less pronounced in amplitude-based reconstructions [12]. End-of-inhale and -exhale phases are defined by a medical doctor (MD) in consultation with a medical physicist (MP). Afterwards, the MD defines the target volumes on the selected phases, which are later used for the definition of the CTV and ITV as per ICRU62. ITV was obtained as a union of the CTVs of the end-of-inhale and exhale phases. Deformable image registration (DIR) is performed between the selected phases using Anatomy Constrained Deformation Algorithm (ANACONDA). Deformation vector fields are evaluated by a MP to quantify the extent of motion within target volume. Motion is assessed in terms of average motion of the target volume and maximum motion of any voxel (point-max). Depending on the extent of the motion the approach to treatment planning and delivery is fine-tuned. For example, decisions are made regarding the field selection and design, enlargement of the spot size [13] exact rescanning strategy [4,6] etc.

### (2) Robust planning

Currently all patients in our clinic receive 3D robust optimized treatments plans. Specifically, for NSCLC patients this decision was based on a preclinical study, in which 3D and 4D robust optimization techniques were compared in terms of achievable plan robustness [14]. This pre-clinical study was conducted utilizing 4D robustness evaluation method (4DREM) introduced by Ribeiro et al. [3]. 3D robust optimization is performed on a single image set (average CT of the planning 4DCT) and accounts for setup and range uncertainty. Optimization for lung cancer patients is performed assuming 6 mm setup uncertainty, while optimization for Hodgkin lymphoma is performed assuming 5 mm setup uncertainty. This is due to immobilization differences and setup reproducibility. Hodgkin lymphoma patients are typically immobilized with a 5-point thermoplastic mask as opposed to NSCLC patients, who are immobilized on a wing board. Estimated range uncertainty for all above mentioned indications is 3%, based on the experimental evaluation as shown by Meijers et al. [15] where the evaluation of range accuracy has been performed for average CT-based calculations, as well as for phase-based calculations. For our patient cohort, all plans incorporated 5-times rescanning (in-layer scaled rescanning). For all lymphoma patients spot size was intentionally enlarged by retracting the range shifter, while for NSCLC patients no intentional spot size enlargement was done. The decision regarding the enlargement of spot size for NSCLC was also based on the pre-clinical study mentioned above, which showed that robust plans can be achieved without enlargement of the spot size, however, during the plan optimization ITV, defined on the average CT, for NSCLC cases was overridden with a muscle tissue density. All lymphoma patients were treated with anterior, anterior-oblique beam arrangement (minimum of 2 fields) and all NSCLC cases were planned with 3-field arrangement. For NSCLC patients anterior, lateral and/or posterior beam directions were used depending on the exact location of the target volume.

Proton spot size in our facility is 3–6.5 mm (sigma) as a function of proton energy, which varies from 230 to 70 MeV respectively. Spot spacing as a function of the spot full width at the half maximum (FWHM) in water with a ratio of 0.8–1 was used during plan optimization. Energy layers were spaced as a function of peak width with a ratio of 0.8–1. Peak width in our facility is 8.7–1.7 mm as a function of energy (230–70 MeV respectively). Layer

switching time is approximately 0.7–0.9 s depending on the energy step.

### (3) Robustness evaluation

Robustness evaluation was performed to assess the outcome of robust optimization. Robustness evaluation was performed in 3D, simulating a set of scenarios with pre-defined setup and range errors. 28 scenarios are calculated, simulating range errors of  $\pm 3\%$  in combination with setup errors of  $\pm 6$  or  $\pm 5$  mm, depending on the indication as mentioned above. Robustness of the plan was assessed on the basis of voxel-wise minimum and voxel-wise maximum dose distributions. In our facility plans are considered robust if V95 of the target volume on voxel-wise minimum distribution exceeds 98% [16].

### (4) Retrospective 4D dose reconstruction

4D dose reconstruction is performed following the method described by Meijers et al. [10]. This method makes use of treatment delivery log files, patient's breathing signals and most recent available 4DCT information throughout the treatment course.

Fraction-wise breathing patterns of each patient are acquired using the Anzai belt system (Anzai Medical, Tokyo, Japan). After the delivery of each fraction, treatment delivery log files are collected. Among other data, log files contain information for every delivered spot regarding its position, dose in terms of monitor units (MU), energy and timing. Based on the timing information

of the breathing signal, delivered spots, as retrieved from the log files, are sorted into corresponding breathing phases. Afterwards, the spots are written into a set of DICOM sub-plans, where each sub-plan contains only the spots associated with a specific breathing phase. These sub-plans are imported into the treatment planning system (TPS), and each sub-plan is calculated on the corresponding phase of the 4DCT. In this step, the most recent available 4DCT is used. In our current clinical practice, patients that are subject to respiratory motion receive weekly repeated 4DCTs. An exception are Hodgkin lymphoma patients, for whom 4DCT in the last week of the treatment may be skipped, if no observations with clinical consequences were made based on daily CBCTs and previous 4DCTs.

Deformation vector fields between a fraction-wise reference phase and all other phases of the 4DCT data set are defined. In the majority of cases, the reference phase is the end-of-exhale phase. Deformation vector fields are used to warp dose contributions of all sub-plans per fraction to the fraction-wise reference phase, where they are summed. Afterwards, deformation vector fields between a course-wise reference phase and fraction-wise reference phase are defined and the fraction-wise summed dose is warped to the course-wise reference phase. On the course-wise reference phase, individual fraction doses are accumulated. The same course-wise reference phase (and target volume) is used throughout the treatment course (also in case plan adaptations have been made). Course-wise reference phase, for example, could be the end-of-exhale phase of the planning CT. Schematically the workflow is shown in Fig. 1.

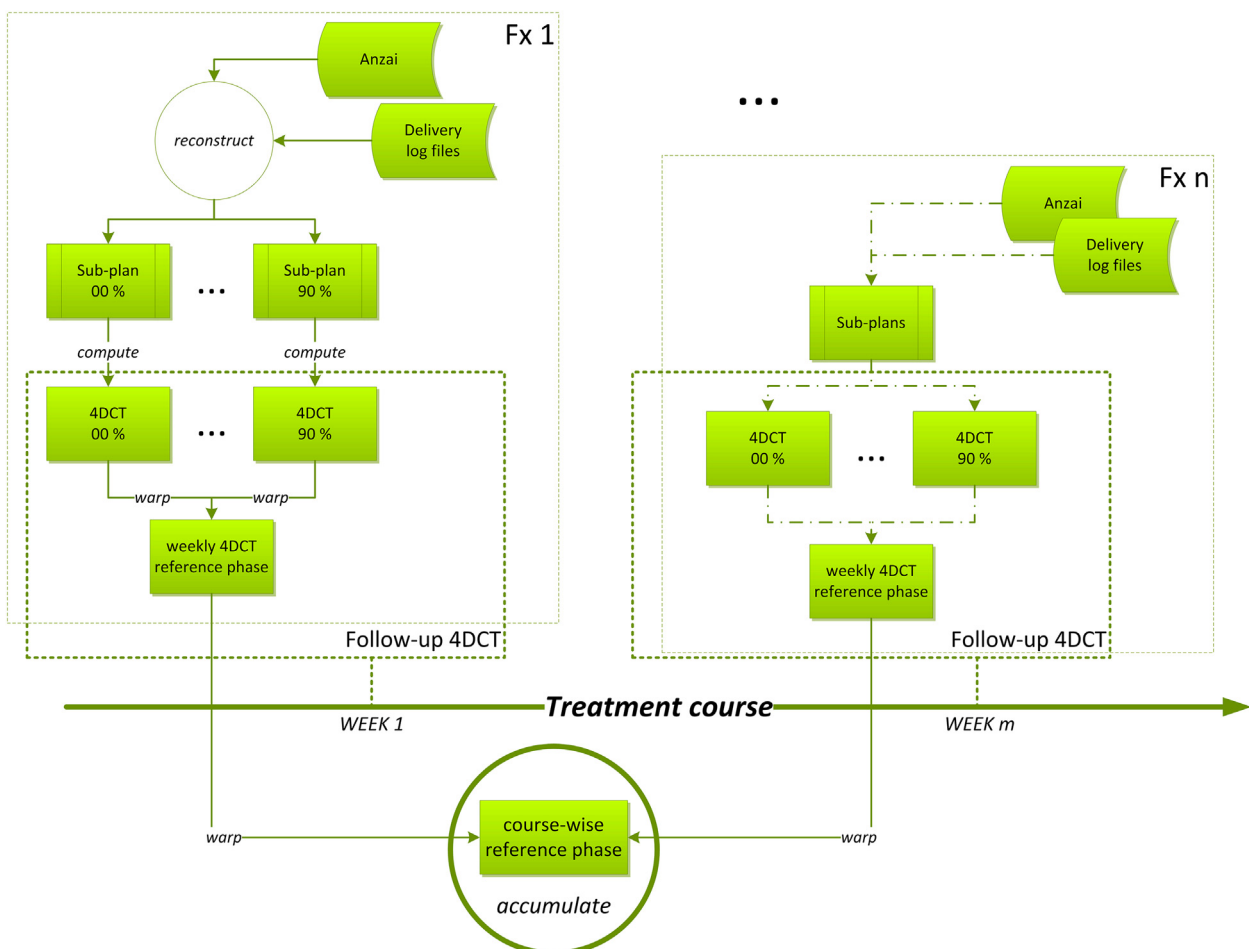


Fig. 1. Schematic representation of the 4D dose reconstruction and accumulation workflow for  $n$  fractions.

Reconstructed fraction doses and accumulated course doses are calculated for the 10 consecutive patients affected by motion and treated at our proton facility. Table 1 summarizes some of the planning and target characteristics of these cases.

ITV V95<sub>vox-min</sub> is V95 based on the voxel-wise minimum dose distribution, derived in the process of robustness evaluation. This parameter is a measure of the robustness for the initial treatment plan.

## Results

Due to the difference in indications, various fractionation schemes and patient-related events, the available data per patient and some treatment characteristics varied among patients. The number of acquired repeat CTs, the number of treatment plan adaptations, the number of fractions, for which breathing signal was acquired and the smallest and the largest observed point-max motions between the end-of-inhale and -exhale phases and mean motions as observed on repeat CTs are listed in Table 2.

For some of the fractions, the acquisition of the breathing signal was skipped either due to logistical issues or patient-related issues. For patient 2 one of the fractions was delivered on a linac due to pending evaluation of the repeat CT, where large anatomical variations were observed. For patient 5 the last two fractions were delivered on a linac in a satellite site due to hospitalization of the patient unrelated to the radiotherapy treatment itself.

All plan adaptations were necessary due to anatomical changes. Changes in the volume of postoperative fluid caused adaptations for patients 2 and 3. For patient 6 adaptation was triggered by a tumor shrinkage, which resulted in an unacceptable dose to OARs. Weight gain required a plan adaption for patient 7. While disappearance of pleural effusion caused plan adaptation for patient 9.

For all follow up 4DCTs the motion evaluation was performed according to the previously described methodology. For example, for patient 6, due to tumor shrinkage, during the course maximum

amplitude of the motion increased significantly, reaching 11 mm as opposed to 5 mm observed in the initial planning CT.

Table 3 shows the summary of observations made on the basis of fraction-wise and accumulated treatment course dose distributions. Mean D98 and D2 (with SD) for the CTV are listed per patient for all reconstructed fractions of the treatment course. In addition, D98, D2 and V95 of the accumulated treatment course dose distribution are shown.

As an example, Fig. 2 shows all reconstructed 4D fractions and accumulated course dose for patients 1 and 2. Additionally supplementary material 1 shows all reconstructed and accumulated doses for all patients.

Out of total 221 reconstructed fractions combined for all 10 patients presented in this study dose to the target volume (D98) remained within 5% from the prescription dose, with only 6 fractions being an exception. In no case, accumulated treatment course dose distributions showed major variations from the prescription dose.

Organ at risk (OAR) doses are summarized in supplementary material 2 for illustrative purposes. However, due to the limitations of accumulated doses, as discussed further, and the set scope of the work, OAR doses will not be discussed in detail.

## Discussion

It can be observed that interplay effects and organ motion introduce loss of dose homogeneity in the target volume on fraction basis. Furthermore, this scales with the degree of target motion. One may notice that the loss of homogeneity for Hodgkin lymphoma is smaller than the loss of homogeneity for NSCLC. For the included NSCLC patients, whole target volumes in the lung were mobile, while for lymphoma patients large parts of the treatment volume are relatively immobile, as parts are located cranially with respect to the lung. Therefore, the organ motion on average is affecting this area to a lesser extent.

**Table 1**

Characteristics of the treatment course preparation specifics for the 10 patients. The point-max motion corresponds to the maximum motion observed for any voxel in the target volume based on the planning 4DCT.

Pat. #	Indication	Prescription	ITV volume, cm <sup>3</sup>	Mean motion, mm	Point-max motion, mm	ITV V95 <sub>vox-min</sub> , %
01	Lymphoma	15 × 2.0 Gy <sub>RBE</sub>	166	0.9	<5	98.13
02	NSCLC	25 × 2.4 Gy <sub>RBE</sub>	44	2.2	<6	99.42
03	NSCLC	25 × 2.4 Gy <sub>RBE</sub>	243	0.9	<6	99.77
04	NSCLC	25 × 2.4 Gy <sub>RBE</sub>	131	0.7	<5	99.82
05	NSCLC	25 × 2.4 Gy <sub>RBE</sub>	357	0.7	<7	99.06
06	NSCLC	25 × 2.4 Gy <sub>RBE</sub>	217	1.1	<5	99.74
07	NSCLC	25 × 2.4 Gy <sub>RBE</sub>	298	1.2	<9	98.41
08	NSCLC	25 × 2.4 Gy <sub>RBE</sub>	202	1.4	<8	98.41
09	NSCLC	25 × 2.4 Gy <sub>RBE</sub>	336	0.6	<6	99.49
10	Lymphoma	15 × 2.0 Gy <sub>RBE</sub>	339	2.1	<20	99.67

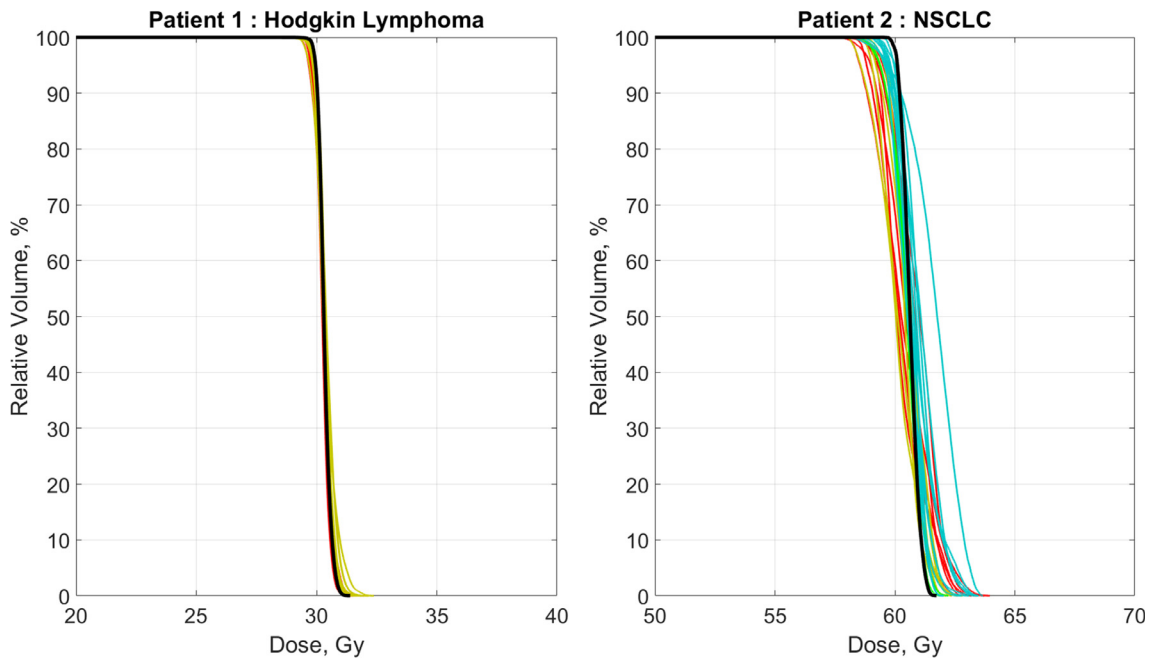
**Table 2**

Available data and treatment characteristics for the 10 considered patients. Mean and point-max motions are shown as the range between the smallest and the largest motion values, as observed on any one specific repeat CT.

Pat. #	Repeat CTs	Plan adaptations	Breathing signals, [available/total]	Mean motion (min–max), mm	Point-max motion (min–max), mm
01	2	0	15/15	(1.1–1.2)	<(6–6)
02	5	2	24/25	(1.8–2.2)	<(7–9)
03	5	1	23/25	(1.0–2.1)	<(7–10)
04	5	0	22/25	(0.8–1.0)	<(5–7)
05	5	0	22/25	(0.7–1.1)	<(7–8)
06	5	1	25/25	(1.8–2.1)	<(7–11)
07	5	1	25/25	(0.5–0.9)	<(7–12)
08	5	0	25/25	(0.9–1.5)	<(7–8)
09	5	1	25/25	(0.7–1.2)	<(5–7)
10	3	0	15/15	(1.1–1.5)	<(11–14)

**Table 3**  
CTV dose statistics based on reconstructed fraction doses and accumulated course dose. Mean D98 and D2 doses and V95 are mean over all the fractions. In the brackets the lowest and the highest values, as observed for any one of the fractions, are provided. A single course-wise CTV, as defined on the reference image set, has been used for the DVH analysis per patient.

Pat. #	Prescription, Gy <sub>RBE</sub>	Mean D98 (min–max), Gy <sub>RBE</sub>	Mean D2 (min–max), Gy <sub>RBE</sub>	Mean V95 (min–max), %	Acc. D98, Gy <sub>RBE</sub>	Acc. D2, Gy <sub>RBE</sub>	Acc. V95, %
01	30	29.7 (29.5–29.8)	31.1 (30.8–31.5)	100 (100–100)	29.9	30.9	100.0
02	60	59.1 (58.3–59.9)	62.2 (61.4–63.3)	100 (100–100)	60.0	61.4	100.0
03	60	57.9 (56.4–58.8)	61.9 (61.4–62.8)	99.1 (96.5–100)	58.4	61.1	99.8
04	60	59.5 (58.9–59.8)	61.6 (61.2–61.9)	100 (100–100)	59.8	61.4	100.0
05	60	59.4 (59.0–59.8)	62.3 (61.6–63.4)	100 (99.9–100)	59.7	62.0	100.0
06	60	58.9 (57.3–59.9)	62.1 (61.3–62.6)	99.9 (99.1–100)	59.8	61.4	100.0
07	60	58.8 (56.4–59.8)	62.3 (61.2–63.1)	99.4 (97.5–100)	59.5	61.9	99.7
08	60	59.2 (58.4–59.7)	61.9 (61.5–62.4)	100 (100–100)	59.7	61.4	100.0
09	60	58.9 (58.0–59.6)	62.0 (61.5–62.6)	99.9 (99.6–100)	59.4	61.6	100.0
10	30	29.6 (29.4–29.7)	30.6 (30.4–31.2)	100 (100–100)	29.7	30.4	100.0



**Fig. 2.** DVHs of CTV for reconstructed fraction-wise 4D dose distributions and accumulated course dose. DVHs are corresponding to cases 1 and 2. Fraction doses are shown in color, while accumulated course dose is shown in black. Fraction doses that have been calculated on the same 4DCT also share the same color. Assigned colors are red, yellow, green, light blue and blue, corresponding from an earlier 4DCT to a more recent in this order.

In the current data set it was not observed that loss of homogeneity induced by motion effects follow a systematic pattern. Systematic patterns generally were caused by anatomical variations, such as, changes in postoperative fluids, patient's weight changes or tumor shrinkage. In all cases, although loss of homogeneity was present on fractional basis, homogeneity was recovered when performing dose accumulation. Local hot and cold spots did not occur in the same location in different fractions.

Although fractionation likely smears out interplay and organ motion effects over the course of radiotherapy, these effects should be considered differently when moving towards hypo-fractionation. In such case, loss of homogeneity very likely may have clinical implications. Nevertheless, it should be noted that for the initial 10 patient data set the observed average target volume motion is relatively low. That is due to the introduced guideline to initially limit the point-max motion below 10 mm during the patient selection (although this was not strictly followed anymore for the patient 10). This illustrates how the 4DREAL methodology can be applied in clinical practice to gradually expand patient inclusion criteria, while ensuring close daily monitoring of the treatment course.

### Limitations

We would like to point out several important assumptions that are made, when performing 4D dose reconstruction and accumulation as described in this study.

- (1) Planning and repeat 4DCTs are assumed to be good representations of patient's anatomy. By reviewing daily CBCTs and comparing them to repeat CTs, one can judge whether the repeat CT is representing the daily anatomy, however this is a subjective evaluation. In case of non-minor inconsistencies, it is difficult to estimate the actual impact of these observations on the calculated dose distributions. In the future, this limitation might be overcome by introducing post-processed synthetic 4DCTs based on daily CBCTs suitable for proton dose calculations. There are multiple examples for developments towards introduction of 4DCBCTs [17] and synthetic CTs [18] which do have improved CT number accuracy that might make synthetic 4DCTs suitable for proton dose calculation. The initial investigations on the

use of artificial intelligence for reduction of motion-induced artifacts in the 4DCT also show promising results [19]. This may further improve 4DCT quality.

- (2) It is assumed that 4DCT, which captures the average patient motion derived from multiple breathing cycles, is representative of the patient's 4D anatomy. Patients exhibiting irregular breathing may be identified calculating the ratio of extreme inhalation amplitude and regular tidal inhalation amplitude [20]. However, breathing cycles are not constant over time. Therefore, potentially a better accuracy of dose reconstruction could be achieved by introducing, so called, "5DCTs". 5DCT can be obtained by combining 4DCTs of variable motion characteristics, each representing an individual breathing cycle. Developments [21] are ongoing, which aim at modeling variations of subsequent breathing cycles. These models can be used to animate 4DCT and generate 5DCTs, incorporating breathing cycle variability. However, validation and, therefore, quantification of accuracy, remains a major challenge for these approaches and eventually these images would provide only an estimation about the motion. Consequently, the added value of 5DCT in dose reconstruction and the impact of its uncertainty remains a topic for further investigations.
- (3) Our dose reconstruction method heavily depends on dose warping. The accuracy and physical meaning of the warped doses is a topic for further investigations. For example, the effect of voxel volume deformation on the meaning of dose must be further clarified. Also, the radiobiological effect on addition of fractionated and varying doses per voxel should be further investigated. For variations of up to +/- 10% in high dose area the additional radiobiological effect has been estimated to be minimal [22]. However, variations per voxel in low dose areas might be much larger and their radiobiological consequences for dose addition is a subject for further clarifications. Due to this, within the scope of current evaluation, we did not investigate low dose areas or dose to OARs, but primarily focused on high dose areas and uniformity of the dose within target volume. To some extent geometric accuracy of the dose warping in the phantom study was investigated during the development of the methodology [10] by being able to reconstruct with a sub-millimeter accuracy the shadow caused by a moving ball bearing in the beam path. However, it has been shown that the accuracy of dose warping reduces when the magnitude of the deformation increases [23]. Also, further investigation on dosimetric consequences caused by the use of different DIR algorithm are required [24].

Currently the smoothness of deformation vector fields is assessed by visual inspection of the deformation grids and accuracy is assessed by manual review of the mapped contours. In case of major anatomical changes, which cannot be attributed to the deformations, deformable image registration would fail. For none of the cases presented here this was the case. However, such scenario is highly probable in clinical routine, in which case corrective actions would be necessary. Otherwise, the meaning of warped doses would become even more questionable. To some extent such situations might be corrected by manual adjustment of control contours and use of them as control ROIs during DIR.

- (4) Currently our implementation of 4D dose reconstruction method does not allow to account for residual patient setup errors or residual beam delivery discrepancies. However, we use 4D dose reconstruction primarily as a tool for gaining insights into interplay and organ motion effects.

In addition to 4D dose reconstruction we perform robustness evaluation on repeat CTs to assess robustness against residual setup errors and range uncertainties. Combined effects of interplay effect, organ motion, setup errors, range uncertainty and changing anatomy indeed might cause some additional dose perturbations. However, since we do not observe systematic patterns linked to interplay from fraction to fraction, it is likely that these additional perturbations would not have severe effects on treatment course dose due to fractionation.

Due to these assumptions and in an absence of further evaluations, we do not recommend considering reconstructed and accumulated dose distributions as "clinical" doses. Therefore, at this stage we would not use accumulated dose distributions, for example, as a background (in other words, "already delivered") dose to be used in plan optimization in case of plan adaptations. However, it would be an attractive use case for the accumulated doses if some of the limitations would be addressed or proven not relevant. This way accumulated hot or cold spots or unintended dose to OARs could be directly mitigated during the plan adaptation process.

The proposed 4D dose calculation workflow can also be employed prospectively by using simulated breathing patterns and log files from the dry-runs. In such way for cases, when motion amplitude exceeds predefined acceptable levels, a set of simulated fractions can be generated to assess if fraction-wise hot/cold spots have systematic behavior. If this is not the case, accumulated DVH would converge towards steeper curve.

Furthermore, a promising future application of accumulated dose distributions could be its correlation with treatment outcomes. This could clarify the clinical relevance of fraction dose variations and could help to reduce uncertainties in the dose parameters enclosed in TCP and NTCP models.

Eventually, the use of daily treatment related information (delivery log files, imaging data, breathing signals, etc.) could be automatically retrieved and processed in a dose accumulation workflow. By introducing warning and action levels for accumulated doses or even using accumulated dose to track TCP and / or NTCP values, it would be possible to implement a whole new layer of quality control longitudinally throughout the treatment course. By using adaptive loops in this process, it would be possible to ensure and gain more confidence that initial clinical goals are met at the end of the treatment course.

In conclusion, the developed methodology for fraction-wise 4D dose reconstruction was applied to 10 consecutive thoracic patients, subject to respiratory motion. Contrary to findings in prospective simulation studies, we did not observe any clinically relevant loss of target dose homogeneity due to interplay and motion effects. Fraction-wise loss of target dose homogeneity due to interplay and organ motion showed no systematic pattern and smeared out with fractionation. Dose degradation caused by anatomical changes showed to be more severe and caused treatment adaptations in five out of ten patients. Although, warped dose distributions should be interpreted with caution, this study provides more realistic incremental insight into the effect of breathing related organ motion on PBS-PT dose delivery.

## Funding

None.

## Disclosures

University of Groningen, University Medical Centre Groningen, Department of Radiation Oncology has active research agreements with RaySearch, Philips, IBA, Mirada, Orfit.

## Conflict of interest statements

Langendijk JA, MD, Prof, is a consultant for proton therapy equipment provider IBA.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.radonc.2020.07.055>.

## References

- [1] Bert C, Rietzel E. 4D treatment planning for scanned ion beams. *Radiat Oncol* 2007;2:24. <https://doi.org/10.1186/1748-717X-2-24>.
- [2] Macdonald OK, Kruse JJ, Miller JM, Garces YI, Brown PD, Miller RC, Foote RL. Proton beam radiotherapy versus three-dimensional conformal stereotactic body radiotherapy in primary peripheral, early-stage non-small-cell lung carcinoma: a comparative dosimetric analysis. *Int J Radiat Oncol Biol Phys* 2009;75:950–8.
- [3] Ribeiro CO, Meijers A, Korevaar EW, Muijs CT, Both S, Langendijk JA, Knopf A. Comprehensive 4D robustness evaluation for pencil beam scanned proton plans. *Radiother Oncol* 2019;136:185–9. <https://doi.org/10.1016/j.radonc.2019.03.037>. Epub 2019 Apr 20.
- [4] Knopf AC, Hong TS, Lomax A. Scanned proton radiotherapy for mobile targets—the effectiveness of re-scanning in the context of different treatment planning approaches and for different motion characteristics. *Phys Med Biol* 2011;56:7257–71. <https://doi.org/10.1088/0031-9155/56/22/016>. Epub 2011 Oct 28.
- [5] Schätti A, Zakova M, Meer D, Lomax AJ. Experimental verification of motion mitigation of discrete proton spot scanning by re-scanning. *Phys Med Biol* 2013;58:8555–72.
- [6] Bernatowicz K, Lomax AJ, Knopf A. Comparative study of layered and volumetric rescanning for different scanning speeds of proton beam in liver patients. *Phys Med Biol* 2013;58:7905–20. <https://doi.org/10.1088/0031-9155/58/22/7905>. Epub 2013 Oct 29.
- [7] Grassberger C, Dowdell S, Sharp G, Paganetti H. Motion mitigation for lung cancer patients treated with active scanning proton therapy: motion mitigation in scanning proton therapy. *Med Phys* 2015;42:2462–9. <https://doi.org/10.1118/1.4916662>.
- [8] Engwall E, Glimelius L, Hynning E. Effectiveness of different rescanning techniques for scanned proton radiotherapy in lung cancer patients. *Phys Med Biol*. 2018;63(9):095006. doi: 10.1088/1361-6560/aabb7b.
- [9] Albertini F, Matter M, Nenoff L, Zhang Ye, Lomax A. Online daily adaptive proton therapy. *BJR* 2020;93:20190594. <https://doi.org/10.1259/bjr.20190594>.
- [10] Meijers A, Jakobi A, Stützer K, Gutierrez Marmitt G, Both S, Langendijk JA, et al. Log file-based dose reconstruction and accumulation for 4D adaptive pencil beam scanned proton therapy in a clinical treatment planning system: implementation and proof-of-concept. *Med Phys* 2019;46:1140–9.
- [11] Richter D, Saito N, Chaudhri N, Härtig M, Ellerbrock M, Jäkel O, et al. Four-dimensional patient dose reconstruction for scanned ion beam therapy of moving liver tumors. *Int J Radiat Oncol Biol Phys* 2014;89:175–81. <https://doi.org/10.1016/j.ijrobp.2014.01.043>. PMID: 24725700.
- [12] Wink NM, Panknin C, Solberg TD. Phase versus amplitude sorting of 4D-CT data. *J Appl Clin Med Phys* 2006;7:77–85.
- [13] Grassberger C, Dowdell S, Lomax A, Sharp G, Shackelford J, Choi N, et al. Motion interplay as a function of patient parameters and spot size in spot scanning proton therapy for lung cancer. *Int J Radiat Oncol Biol Phys* 2013;86:380–6. <https://doi.org/10.1016/j.ijrobp.2013.01.024>.
- [14] Ribeiro CO, Visser S, Meijers A, Korevaar EW, Wijsman R, Muijs C, et al. Towards the clinical implementation of Pencil Beam Scanned Proton Therapy (PBS-PT) for thoracic indications: a comprehensive 4D robustness evaluation based on patient and machine specific information. *Int J Radiat Oncol Biol Phys* 2019;105:E754–5. <https://doi.org/10.1016/j.ijrobp.2019.06.779>.
- [15] Meijers A, Seller Oria C, Free J, Bondesson D, Rabe M, Parodi K, et al. Assessment of range uncertainty in lung-like tissue using a porcine lung phantom and proton radiography. *Phys Med Biol*. 2020. doi: 10.1088/1361-6560/ab91db.
- [16] Korevaar EW, Habraken SJM, Scandurra D, Kierkels RCJ, Nipan M, Eenink MGC, et al. Practical robustness evaluation in radiotherapy – a photon and proton-proof alternative to PTV-based plan evaluation. *Radiother Oncol* 2019;141:267–74. <https://doi.org/10.1016/j.radonc.2019.08.005>.
- [17] Niepel K, Kamp F, Kurz C, Hansen D, Rit S, Neppel S, et al. Feasibility of 4DCBCT-based proton dose calculation: an ex vivo porcine lung phantom study. *Zeitschrift für Medizinische Physik* 2019;29:249–61. <https://doi.org/10.1016/j.zemedi.2018.10.0>.
- [18] Botas P, Kim J, Winey B, Paganetti H. Online adaption approaches for intensity modulated proton therapy for head and neck patients based on cone beam CTs and Monte Carlo simulations. *Phys Med Biol*. 2018;64(1):015004. doi: 10.1088/1361-6560/aaf30b.
- [19] Werner R, Sentker T, Madesta F, Gauer T, Hofmann C. Intelligent 4D CT sequence scanning (i4DCT): concept and performance evaluation. *Med Phys* 2019;46:3462–74.
- [20] Zeng C, Plastaras JP, Tochner ZA, White BM, Hill-Kayser CE, Hahn SM, Both S. Proton pencil beam scanning for mediastinal lymphoma: the impact of interplay between target motion and beam scanning. *Phys Med Biol* 2015;60:3013–29. <https://doi.org/10.1088/0031-9155/60/7/3013>. Epub 2015 Mar 19. PMID: 25789418.
- [21] Zhang Ye, Knopf A, Tanner C, Boye D, Lomax AJ. Deformable motion reconstruction for scanned proton beam therapy using on-line x-ray imaging. *Phys Med Biol* 2013;58:8621–45. <https://doi.org/10.1088/0031-9155/58/24/8621>.
- [22] Bortfeld T, Paganetti H. The biologic relevance of daily dose variations in adaptive treatment planning. *Int J Radiat Oncol Biol Phys* 2006;65:899–906.
- [23] Spautz S, Peters N, Meijers A, Jakobi A, Knopf AC, Troost EGC et al. Accuracy and robustness of 4D logfile-based dose reconstruction. Conference proceedings 4D workshop 2019, Krakow.
- [24] Ribeiro CO, Knopf A, Langendijk JA, Weber DC, Lomax AJ, Zhang Y. Assessment of dosimetric errors induced by deformable image registration methods in 4D pencil beam scanned proton treatment planning for liver tumours. *Radiother Oncol* 2018;128:174–81. <https://doi.org/10.1016/j.radonc.2018.03.001>. Epub 2018 Mar 20.