Aus der Klinik für Strahlentherapie und Radioonkologie

der Medizinischen Fakultät Mannheim

(Direktor: Prof. Dr. med. Frank A. Giordano)

Evaluation of two adaptive radiotherapy workflows and their benefits for stage III non-small cell lung cancer patients using cone-beam computed tomography-based synthetic computed tomography

Inauguraldissertation

zur Erlangung des Doctor scientiarum humanarum (Dr. sc. hum.)

der

Medizinischen Fakultät Mannheim

der Ruprecht-Karls-Universität

zu

Heidelberg

vorgelegt von

Lea Gesa Hoppen

aus

Nordhorn

2023

Dekan: Prof. Dr. med. Sergij Goerdt

Referent: Prof. Dr. med. Frank A. Giordano

Table of Contents

List of abbreviations

AD	Adenocarcinoma
AI	Artificial intelligence
AJCC	American Joint Committee on Cancer
ART	Adaptive radiotherapy
CBCT	Cone-beam computed tomography
cCRT	concurrent chemoradiotherapy
CT	Computed tomography
CT-to-ED	CT-number-to-electron density
CTV	Clinical target volume
DD	Dose difference
DIBH	Deep inspiration breath-hold
DIR	Deformable image registration
DL	Deep learning
d_{mean}	Mean surface distance
DP	Difference dose plans
D_{Presc}	Prescription dose
DSC	Dice similarity coefficient
DTA	Distance-to-agreement
DVF	Deformation vector field
DVH	Dose volume histogram
$D_{x\%}$	Maximal dose a relative volume receives
ED	Electron density
EUD	Equivalent uniform dose
EQD_2	Equivalent dose in 2Gy-fractions
f	female

FOV	Field of view
fx	Fraction
GAN	Generative adversarial network
GPR	Gamma pass rate
GTV	Gross tumor volume
HD	Hausdorff distance
HU	Hounsfield units
IGRT	Image guided radiotherapy
IMRT	Intensity modulated radiotherapy
IQR	Interquartile range
IRB	Institutional review board
ITV	Internal target volume
IVH	Intensity-volume-histogram
L0	0-1 loss function
L1	mean absolute error loss function
L2	mean squared error loss function
LLL	Left lower lobe
LML	Left middle lobe
LUL	Left upper lobe
m	male
MD	Microscopic disease
MLC	Multi-leaf collimator
MAE	Mean absolute error
ME	Mean error
MED	Mean esophagus dose
MHD	Mean heart dose
MLD	Mean lung dose

MR	Magnetic resonance
MRI	Magnetic resonance imaging
NSCLC	Non-small cell lung cancer
NTCP	Normal tissue complication probability
OAR	Organs at risk
OS	Overall survival
P#	Patient number (1-15 & 1-13)
PCC	Pearson's correlation coefficients
pCT	planning computed tomography
PET-CT	Positron emission tomography – computed tomography
PFS	Progression-free survival
PTV	Planning target volume
PTV_1	initial PTV on the first sCT
QA	Quality assurance
QO	Quadratic overdose
QU	Quadratic underdose
Q-Q plot	Quantile-quantile plot
RLL	Right lower lobe
RML	Right middle lobe
RT	Radiotherapy
RUP	Right upper lobe
SAM	Segment aperture morphing
SCC	Squamous cell carcinoma
SCLC	small cell lung cancer
sCT	synthetic computed tomography
sCT_1	synthetic computed tomography of the first treatment fraction
sCT_{2-n}	all sCTs of one patient except sCT_1

SD	Standard deviation
SEN	Sensitivity
SPEC	Specificity
SWO	Segment weight optimization
TCP	Tumor control probability
TP	Treatment plan
TNM	Tumor, nodes, and metastases
VOI	Volumes of interest
V_{xGy}	Volume receiving x Gy
VMAT	Volumetric modulated arc therapy
WHO	World Health Organization
XVI	X-ray volume imaging

1. Introduction

Cancer is one of the leading causes of death worldwide, and of these, patients with lung cancer have the highest mortality rate, accounting for approximately 21% of all cancer deaths [1]. Surgery is the most effective treatment for patients with lung cancer, but most patients with this condition are considered inoperable [2]. Approximately 76% of all lung tumors are histologically diagnosed as non-small cell lung cancer (NSCLC) [3]. The gold standard for treating NSCLC patients with a good performance status is concurrent chemoradiotherapy (cCRT) [4]. The range of a sufficient tumor control probability (TCP) with an acceptable normal tissue complication probability (NTCP) is defined as the therapeutic window. To broaden the therapeutic window, the goal of radiotherapy is to deliver sufficiently high doses to the target to obtain a high TCP, while minimizing the doses to the organs at risk (OARs) to achieve a good NTCP [5]. The main limiting factor for patients with lung cancer are the maximum tolerated doses to the lungs, heart, and esophagus. Patients whose tolerance doses of these OARs are exceeded have a greater tendency to develop pneumonitis [2, 6, 7], cardiovascular toxicity [8-11], and acute esophagitis [12, 13], respectively. With an approximate incidence of 30 %, radiation induced pneumonitis remains the most severe dose limiting parameter in the radiotherapy of patients with NSCLC [14]. The most common fractionation scheme for patients with NSCLC is 60 Gy-70 Gy in 1.8 Gy-2 Gy-fractions [9, 15-18].

Radiotherapy has continuously improved in terms of dosimetric accuracy over the last decades. The introduction of intensity modulated radiotherapy (IMRT) enables radiation treatment plans to consist of multiple overlapping sub segments each formed by a multi-leaf collimator (MLC) with defined individual dose contributions. A further development of IMRT is volumetric modulated arc therapy (VMAT). VMAT delivers a dose with precise coordination of the MLC movement and gantry rotation speed and enables the MLC to form the sub segments while the gantry continuously rotates around the patient [19]. These techniques allow precise adjustments of the dose distributions even for complex target volumes and provide good protection of the surrounding OARs. Furthermore, with image guided radiotherapy (IGRT) the translational and rotational patient positioning is monitored with daily cone-beam

computed tomographies (CBCTs), matched with the corresponding treatment planning computed tomography (pCT) and adjusted using couch-shift. Despite these improvements, the 5-year overall survival rate (OS) for patients with NSCLC undergoing definitive cCRT remains low at 32 %-33 % [20, 21]. Subsequent immunotherapy with durvalumab for up to 12 months for patients with a World Health Organization (WHO) performance status < 1 and no toxic events ≥ grade 2 improved the 5-year-OS to 43 % [22, 23].

In the current workflow, treatment plans are simulated on a single computed tomography (CT) scan obtained prior to the treatment. These treatment plans are typically delivered in all treatment fractions using daily IGRT. To account for respiratory motion, IGRT is often combined with a motion management approach. For the passive motion management, intrafractional tumor motion is usually addressed by an internal target volume (ITV) margin [24, 25]. Active motion management takes tumor motion into account through beam tracking or gating concepts [26, 27]. The ITV margin includes the planning target volume (PTV) in each breathing phase so that the tumor is adequately covered throughout the whole respiratory cycle. The disadvantage of the ITV margin is that it enlarges the target volume and thereby simultaneously also the high dose region of the normal tissue [27]. Real-time tracking can be performed either by following the tumor movement with the treatment beam or by moving the patients with the help of a robotic couch [25]. Real-time target tracking often uses metal fiducial markers placed close to the target, allowing for adjusted irradiation [28]. When gating techniques are used, the target volume is only irradiated during a specific phase of the respiratory cycle [24, 27]. Here, in contrast to tracking concepts, the movement of the target volumes were reduced to a minimum. Beam tracking and gating concepts address a quasi-static scenario and decrease the PTV size. These procedures ensure appropriate patient and, by extension, precise tumor positioning based on the initial treatment plan. However, they do not account for interfractional morphologic changes in the anatomy of the patient due to weight loss, lung volume, tumor position, or tumor size during the treatment period. These changes could lead to discrepancies between the planned and delivered dose [5, 29, 30]. This is especially relevant for patients with NSCLC, as these tumors generally respond early to radiation. Therefore, they often begin to shrink early [30, 31]. On average, the gross tumor volume (GTV) of patients with NSCLC decreases by 0.6 %-2.4 % per day during radiotherapy [2, 6, 32].

Based on the decreasing target volume additional improvement of the treatment quality could be achieved by implementing adaptive radiotherapy (ART) workflow strategies. To minimize the discrepancies between the planned and delivered dose and, therefore, further improve the therapy for patients with lung cancer, in ART, treatment plans are adapted to the patient's potentially altered anatomy [5, 29, 30], based on daily pretreatment images. With the use of ART, it is also possible to reduce safety margins due to lower uncertainties regarding the location and shape of the tumor, which would further intensify the positive effect of ART [2, 33]. Treatment plan adaptations can be performed at three different time points: offline in between treatment fractions, online immediately prior to a treatment fraction, or in real-time during the treatment fractions [34].

By adapting the treatment plan to the actual patient anatomy, the OAR-dose can be reduced without compromising the PTV coverage (isoeffective), the dose to the target volume can be further escalated without increasing ipsilateral lung-/OAR-dose (isotoxic) or a compromise of the two scenarios can be chosen [30]. In both scenarios, the therapeutic window would be widened. By minimizing doses to the OAR, the occurrence probability of toxic events can be decreased and the severity in case of toxic events can be decreased. The V_{20Gy} (the relative volume of a specific organ that receives more than 20 Gy) and MLD of the ipsilateral lung correlate with the dose-limiting toxicity pneumonitis [2]. Cardiac toxicities are associated with higher V_{5Gy}(heart) and mean heart dose (MHD) [35]. Patients with a higher mean esophagus dose (MED) are more likely to experience an acute esophagitis [12, 13]. Furthermore, lower heart and lung doses correlate with an improved OS [8, 9, 36]. The double-blinded PACIFIC trial compared a subsequent immunotherapy group with durvalumab, which can further improve the survival rates, with a placebo group. This study included patients with a WHO performance of 0-1 and the absence of toxic events ≥ grade 2 [23]. Despite the increase of the survival rate, higher incidence rates of pneumonitis occurred in the durvalumab group [37]. It is beneficial to minimize the doses to the OAR and thus the risk of pneumonitis and other toxic events ≥ grade 2 to give the patients a better quality of life and increase the possibility of a subsequent immunotherapy [11, 33].

Several steps are necessary to implement an ART workflow. First, daily calibrated images are required on which one can accurately calculate the dose distribution. Second, adjusted structures of the OAR, GTV and PTV are needed within the shortest possible time. Third, the generation of the adapted treatment plans should be accurate and fast. Fourth, the dose distributions of the individual treatment fractions must be accumulated in order to analyze the total dose. In addition, each workflow step requires a quality assurance (QA).

Usually, a CBCT is acquired at each treatment fraction. When using these CBCTs for ART, the CBCTs' insufficient image quality is one of the major challenges, which needs to be overcome [38-40]. Compared to fan-beam CTs, CBCTs have severe image artifacts due to beam hardening, detector scatter, image lag, and patient-specific scatter, thus there is no unique CT number-to-electron density (CT-to-ED) calibration, which makes dose calculations more error-prone. Currently, many calibration methods already exist to overcome this problem for accurate dose calculations on CBCTs [41]. Among these are hardware-based preprocessing artifact correction approaches [42], bulk density override [43], patient or population specific calibration curves (CT-to-ED assignment) [44], deformable image registration (DIR) [38, 39], and image processing algorithms that further improve the image quality [45-47]. The use of artificial neural networks to generate synthetic CTs (sCTs) has become increasingly popular over the last several years. They can be trained with paired multi-modal datasets (for example, CBCT and CT) and preserve the actual anatomy of the daily images in the resulting sCTs. Several studies have already shown promising results in terms of artifact reduction and thus for the potential of a clinical implementation of an adaptive workflow [48-53].

Manual contouring of relevant organs and target volumes is time-consuming. For ART, a fast image segmentation is required. DIR and deep learning (DL) algorithms are lately often used for the automated image segmentation. A DIR between the reference and object images creates a displacement vector for each voxel, thereby forming a deformation vector field (DVF) to deform the structures on the object image [54]. The DVF generation is strongly affected by the image resolution, similarity of the organs, the image quality, and image distortion [55-57]. Furthermore, the generation of the DVFs is a problem with no unique solution due to too many degrees of freedom.

Therefore, the results of different DIR algorithms are often differing [58]. In the atlas-based method, pre-contoured structures of the atlas templates are deformed to the object images. Image segmentation accuracy can be improved by using more atlas datasets. However, a DIR is very computationally complex, which further proliferates as the atlas template pool becomes larger [59]. During the last years, DL-based automatic image segmentation has been used more and more with fully convolutional network architectures, which allows end-to-end training and pixel-to-pixel segmentation. Therefore, they are able to adapt to customized configurations of contouring procedures and image modalities of a department [59]. In an auto segmentation challenge, the DL algorithms performed well and performed better than the methods based on multi-atlas [60].

The time required for the adaptations of the treatment plans should be minimized without affecting the quality of the generated treatment plans. Fast treatment plan preparation for ART can be achieved by several methods. The use of a library with a collection of previously created treatment plans enables to choose the most appropriate "plan of the day" for each adaptation [5, 61]. Alternatively, the initial treatment plan can be modified by using segment aperture morphing (SAM) and/or segment weight optimization (SWO). SAM calculates the needed MLC morphing to adapt the segments to the current target contours based on spatial relationships between the initial target and the current target contours [5, 62]. SWO optimizes the weight of each segment [62].

Dose accumulation of the adapted treatment plans is usually performed by using the inverse DVFs produced through a DIR. This method is only as accurate as the DIR, which, as mentioned earlier, depends to a large extent on the image artifacts, image resolution and image distortion [55, 56]. In addition, due to too many degrees of freedom, the DVF generation is a problem that has no definite solution [58]. This may lead to dose inaccuracies in the DIR-based dose accumulation [56]. This inaccuracy is even more distinct for regressive structures since their mass is not preserved. This makes dose accumulation more difficult in the case that shrinking tumors are included [63, 64].

Finally, QA of each workflow step of ART is necessary to ensure the safe and accurate implementation of ART [65]. Various QA techniques already exist for this purpose: Analytical methods for the accuracy of image registration and segmentation [65, 66], measurements with a deformable phantom [67], programs for a quality and reliability check of the treatment plan [65, 68], dose distribution validation of the treatment planning system by using second dose engines [65, 69, 70], real-time dose delivery tracking with portal imaging devices or transmission detectors, and a dose reconstruction using log files that provides information for each segment of dose rate, leaf positions of the MLC, and gantry angles [5, 65, 71]. Cai et al. have implemented a QA strategy for these steps that is not time consuming [65].

Regarding dose escalation, several studies indicated that for patients with NSCLC higher fractional doses resulted in higher TCP and OS [72-74]. This was questioned by the outcome of the RTOG-0617 study. RTOG-0617 compared the results of normofractionated patients (60 Gy in 2 Gy fractions) with a high-dose group (74 Gy in 2 Gy fractions). Surprisingly, patients treated with 60 Gy had a better OS and PFS than those treated with 74 Gy [9]. However, the lungs, heart, and the esophagus received higher doses in the high dose patient group than in the normofractionated patient group [9, 36]. Higher heart and lung doses are associated with a worse OS [8, 9, 36, 75]. Further analysis showed that dose in a particular subregion at the heart base that includes the left coronal artery origin and is located near the sinoatrial node, was significantly associated with OS [76]. This might possibly explain the poorer outcome of the high-dose group [76]. Furthermore, re-analysis of the RTOG-0617 patient population revealed that the dose escalation improved the outcomes for patients with a radioresistant genotype [77]. Moreover, in this study, the number of treatment fractions was increased rather than the fractional dose, resulting in a prolonged treatment period [9, 75, 78]. Although RTOG-0617 questions whether conventional dose escalation is beneficial, different fractionation schemes for dose escalation could yield better outcomes than conventional radiotherapy (RT) [79]. Therefore, a target dose escalation with higher fractional doses without exceeding the doses to the adjacent OARs might be beneficial especially for patients with a radioresistant genotype.

One concern regarding ART is the under dosage of the microscopic disease (MD) [30]. It is unclear if the MD shrinks synchronously with the tumor regression or remains stationary [5]. In a study by Guckenberger et al. with two treatment plan adaptations, no under dosage of the MD occurred with a fractional dose escalation and a constant MLD [32]. Hence, by adapting the treatment plan to the shrinking GTV, it might be possible to escalate the fractional dose without increasing the doses to the OARs while preventing MD underdosing.

For widespread clinical use of ART, the additional effort or rather the additional time needed should be minimal, as little different software as possible should be needed, there need to be standard guidelines especially regarding quality assurance, and it should be clarified which patients benefit most from ART and how often adjustments of the treatment plan are needed. The uncertainty of MD underdosing in isoeffective ART and the controversial dose escalation in isotoxic ART complicate clinical implementation in patients with NSCLC.

The aim of this work is to evaluate the dosimetric benefit of daily isoeffective and isotoxic ART for patients with stage III NSCLC using CBCT-based sCTs generated using a trained artificial intelligence (AI) algorithm. For this purpose, this work first focuses on the generation of these sCTs using a dedicated trained AI algorithm and subsequently analyzes the accuracy of the sCTs in terms of image quality and dosimetric accuracy. Subsequently, the accuracy of a DIR algorithm deforming the manually delineated structures from the pCT to the sCT is evaluated. At the end, treatment plans for the isoeffective and isotoxic ART scenarios were generated, compared with the actual delivered and initially planned dose, and evaluated for their dosimetric benefit.

The most relevant results of this work have already been published in two research papers [80, 81].

2. Materials and methods

The following four subprojects, as also illustrated in Fig. 1, were performed sequentially to facilitate the introduction of isoeffective and isotoxic ART for NSCLC patients.

1. As a first step the problem of insufficient CBCT image quality and consequently a missing CT-to-ED calibration curve caused by various image artifacts needed to be overcome. For this purpose, multimodal paired image data (1 CT, 5 CBCTs per patient) of the thoracic region were collected and provided to Elekta. They trained an AI-based algorithm (2D cycle-generative adversarial network (cycle-GAN)) for generating sCTs out of CBCTs without the need of a pCT. As a result, a specific sCT model for this anatomical region was developed [80]. The first part of this thesis briefly reviews the architecture of the applied cycle-GAN, which is included in the research software ADMIRE AI (Elekta AB).

2. Since an accurate assignment of the gray values to the electron density is necessary to ensure precise dose calculations on the sCT, the image quality of the generated sCTs was checked in the second step. For this purpose, the generated sCTs of the first fractions (sCT_1) were compared with the pCT with regard to gray values and dosimetric deviations.

3. Since the time factor also plays a major role in ART, it would be very beneficial to obtain accurate automatically generated structures. For this reason, in the third subproject, a DIR algorithm in ADMIRE AI was tested. Structures were deformed from the pCT to the sCT_1. The deformed structures were compared with manually delineated structures using the analysis parameters Dice similarity coefficient (DSC), mean surface distance (d_{mean}), Hausdorff distance (HD), specificity and sensitivity.

4. Finally the initial treatment plan was retrospectively recalculated (IGRT) and two additional treatment plans were generated on the remaining fractions of each patient (sCT2-n), one with an escalated PTV dose (isotoxic ART) and one with reduced OAR-doses (isoeffective ART). They were compared with the initial treatment plan to analyze the dosimetric benefit of isotoxic and isoeffective ART scenarios. The dosimetric advantages of ART techniques over current conventional IGRT were quantified by comparing several dose volume metrics.

2.1 Generation of sCT model

| 53 patient data sets (1CT + 5CBCTs) | → | Training of 2D cycle-GAN | → | Specific thoracic sCT model |

2.2 Evaluation of image quality of sCTs

| Difference images & IVH comparison | **+** | ME & MAE compared to pCT | **+** | Dosimteric Gamma-analysis |

2.3 Accuracy of deformed structures

| Deformed structures from pCT to sCT_1 | **vs.** | Manually delineated structures | → | DSC, HD, d_{mean}, SEN, SPEC |

2.4 Benefit evaluation of daily offline ART in Monaco

| Daily CBCTs | → | Generation of sCT | → | Corrected deformed structures | → | Plan adaptation procedures on sCTs |

| → | Initial TP | **vs.** | • IGRT • isoeffective ART • isotoxic ART | → | Dosimetric comparison |

Figure 1 - The four subprojects which were performed and analyzed successively. First the training of the cycle-generative adversarial network (cycle-GAN) for the synthetic CT (sCT) model. Second, the evaluation of the sCT's image quality by evaluating the difference images and the intensity volume histograms (IVH) in comparison to the initial treatment plan (TP) and calculating the mean error (ME), mean absolute error (MAE), and the gamma pass rate. Third, the accuracy of the deformed structures from the planning CT (pCT) to the sCT of the first treatment fraction (sCT_1) compared to manually delineated structures in terms of the dice similarity coefficient (DSC), Hausdorff distance (HD), mean surface distance (d_{mean}), sensitivity (SEN), and specificity (SPEC). Finally, to evaluate the benefit of daily offline adaptive radiotherapy (ART) for patients with non-small cell lung cancer stage III regarding organ at risk sparing and target dose escalation compared with image-guided radiotherapy (IGRT) without adaptation.

2.1 sCT generation for the thoracic region

In order to ensure accurate dose calculation for ART, it is necessary to compensate for the insufficient image quality. In many medical fields, AI and DL are used with increasing frequency and are becoming more and more important. For example, images can be processed and artifacts such as scatter artifacts and image noise can be reduced or, in the best case, completely eliminated [82]. For image synthesis of unpaired datasets, generative adversarial networks (GAN) have shown superior results regarding their image quality and accurate shape preservation [83-86]. The shape preservation is of great importance for patient images so that the anatomy is not distorted. In a forward GAN the generator has been trained to make the images to be processed ($images_{process}$) to look like the goal images ($images_{goal)}$. In a feedback loop the discriminator compares the synthetically created image ($image_{synthetic}$) with the $images_{goal}$ and tries to determine with a loss function whether the $image_{synthetic}$ belongs to the $images_{goal}$ or whether it is a synthetically created one [82, 83, 86, 87]. Combining a forward with a backward GAN, which tries to process $images_{goal}$ to look like $images_{process}$, can further improve the image quality and minimize discrepancies between the $images_{process}$ and $images_{goal}$. This is the case in cycle-GAN [87]. In a cycle-GAN, different loss functions are used to penalize differences between $images_{process}$ and $images_{goal}$. Commonly used loss functions are, for example, L0, L1 and L2. L0 (0-1 loss) penalizes all deviations to $images_{goal}$ (or all deviations above a certain value) equally strong and thus prefers that the most pixels of two images are identical, but permits larger deviations in the rest of the pixels. L1 (mean absolute error loss) uses the absolute error values for punishment. L2 (mean squared error loss) uses the squared values for penalizing. Therefore, it strongly penalizes large deviations, but allows smaller deviations in large areas [83]. In this work, an AI algorithm has been used to generate sCTs that can be used for dose calculation for ART. Hence, these sCTs are a potential tool for ART. This algorithm was implemented in the research software ADMIRE AI.

2.1.1 The AI model in ADMIRE AI

For this work the research software ADMIRE AI was used to generate sCTs out of daily CBCTs. In a research software of the treatment planning system Monaco, ADMIRE AI is implemented as a standalone module. In ADMIRE AI a 2D cycle-GAN framework enables ADMIRE AI to produce sCTs based on CBCTs. On the basis of the network structures, generators, and discriminators presented by Zhu et al. [88] and Xu et al. [89], Elekta further developed the cycle-GAN framework using data provided by our research group to enable a CBCT application.

The main problems with applying the original cycle-GAN to medical image syntheses are possible geometry or structure distortion of the generated images and that the synthesized images may not obtain pixel-level accuracy due to pure distribution matching in two domains [90]. However, for our purpose, it is crucial to strictly preserve the original patients' anatomy of the CBCTs and achieve an accurate pixel-level with CT numbers in the generated sCT images. Therefore, some modifications were made to overcome the above-mentioned problems of the original cycle-GAN approach. First, unlike the unpaired training strategies used in many previous research papers, the training mechanism in this sCT model is based on paired datasets of CBCT and pCT images [50, 91, 92]. Second, in addition to the original loss and cycle loss terms in the original cycle-GAN framework, an extra L1 loss function was added between the synthetic images and the corresponding CT images for the forward and backward direction. Third, to further prevent possible structural distortions, some transformation layers were added to the initial networks. With these alterations, the trained sCT module can preserve the original CBCT structures in the generated sCTs and also achieve the pixel-level accuracy of CT numbers [80].

In order to achieve better standardization and shorter training times for improved results, some data pre-processing was done before training [80]. All datasets were trimmed into the CT number range of -1000 Hounsfield units (HU) to 4000 HU to better represent high density structures. For the training processes, all images were scaled linearly to [-1.0 1.0] as the input and the predicted output. Additionally, all images have been resampled to the size of 224 pixels $\times$ 224 pixels.

2.1.2 Patient population and image acquisition for the sCT-model training

We collected paired pCT and CBCT data of the thoracic region for 53 patients to train the cycle-GAN algorithm. Each patient dataset consisted of one pCT and the first five CBCTs. All pCT images were acquired using a Brilliance Big Bore CT scanner (Philips Healthcare, Amsterdam, Netherlands). The CBCT acquisitions were made daily prior to each treatment fraction with a kV-based X-ray volume imaging (XVI) scanner (Elekta AB). The CT-to-ED calibration curve and the standard deviation (SD) of the pCT's image noise were obtained using a tissue-characterization phantom (model 467, Gammex, Middleton, WI, USA). The imaging parameters for the pCT/CBCT generation were as follows: a tube voltage of 120 kV/120 kV, an exposure time product of 116 mAs/132 mAs, slice sickness of 3 mm/3 mm, and the standard B/F1 Bow-tie filter. The collimator position for the CBCT was M20 which yielded a field of view (FOV) of 27.6 cm*42.6 cm. These CBCT scans were exported after reconstruction using the clinically applied rigid registration, where a match to the tumor was preferred. All patients were treated in 2018 or 2019. The datasets for the training of the cycle-GAN algorithm were entirely different from the datasets, which were used to test the image quality of the sCTs in order to avoid overfitting the results. The patient characteristics are displayed in Tab. 1.

All analysis in this work was conducted after Institutional Review Board (IRB) approval (2018-836R-MA) and in accordance with the principles of the Declaration of Helsinki. All data were anonymized before inclusion.

Table 1 - Patient population for the training of the synthetic CT (sCT)-model. Listed are the diagnoses, the tumor stages, the total and fractional prescription doses, and the average patient age (range) comprised for all patients of the patient population.

Patients	Diagnosis	Tumor stage	Prescription dose (Gy)	Fraction dose (Gy)	Mean patient age (range) (years)
53	lung cancer, intrathoracic lymph node cancer	T2b-T4a	60,66,70	2	66.1 (51-90)

2.2 Evaluation of the generated thoracic sCT

To use the generated sCTs for dosimetric calculations first the image quality must be evaluated. For this purpose sCTs of the first fraction CBCTs ($CBCTs_1$) of an entirely different patient cohort than the training patients were compared with their corresponding pCTs. The $CBCTs_1$ were used to minimize the anatomical differences between the generated sCT and the pCT.

2.2.1 Patient population

For the image quality evaluation 15 datasets of patients with stage II-IV NSCLC were evaluated. Tumors were staged using the American Joint Committee on Cancer (AJCC) tumor, nodes, and metastases (TNM) staging system version 8. One dataset consisted of one pCT, the $CBCT_1$, and sCT_1. The characteristics of the patients are displayed in Tab. 2. In analogy to the datasets for the cycle-GAN training the pCT and CBCT images were obtained with the Brilliance Big Bore CT scanner and the daily kV-based XVI scanner prior to each treatment fraction, respectively. The imaging parameters are the same as those of the training datasets.

2.2.2 Analysis of image uncertainties

For a potential adaptive workflow, precise dose calculations require an accurate CT-to-ED calibration. An accurate CT-to-ED calibration is achieved by sufficient image quality through artifact removal. The image quality can be evaluated using various parameters. These include the commonly used parameters, mean error (ME) and mean absolute error (MAE), which were chosen for the analysis in this work [93, 94].

First the sCTs were visually checked for structural changes with respect to the CBCT. Than the image quality of the CBCT and sCT was compared by generating difference images (pCT-CBCT and pCT-sCT) after matching the CBCT and sCT to the pCT. For this analysis the images were imported into the software Velocity (version 3.2.1, Varian Medical Systems, Palo Alto, USA). Furthermore, two volumes of interest (VOIs) were generated using the manually delineated structures of the patient outline and lungs. Margins of 10 mm and 15 mm for potential anatomical differences between the two

images were subtracted from the lung and the patient outline to obtain the lung VOI and total body VOI, respectively. An example for the two VOIs is depicted in Fig. 2. The negative margin for the VOIs were only used for the intensity-volume histogram (IVH), ME, and MAE analysis.

Table 2 - Patient population for the evaluation of the synthetic CTs (sCTs) generated by the trained cycle-GAN algorithm and the image segmentation accuracy analysis. Listed are the patient's sex, age, tumor staging, carcinoma subtype, tumor location and the total prescription dose (D_{presc}) with the number of fractions (fx) for each patient. For the age the average age and the range was listed additionally.

Patient	Sex	Age	Tumor staging				Carcinoma subtype	Tumor location	D_{presc} (Gy)/fx
			T	N	M	Stage			
P1.1	m	88	3	0	0	IIB	SCC	LLL	70/35
P1.2	m	61	3	2	1	IV	SCC	LUL	66/33
P1.3	m	69	4	0	0	IIIA	SCC	RUL	70/35
P1.4	f	55	4	3	0	IIIC	AC	RUL	60/30
P1.5	f	65	4	3	0	IIIC	AC	RLL	60/30
P1.6	f	66	2b	0	0	IIA	SCC	LUL	70/35
P1.7	m	70	4	2	0	IIIB	AC	RUL	66/33
P1.8	f	49	4	1	0	IIIA	SCC	RLL	54/27
P1.9	m	49	3	3	0	IIIC	SCC	LUL	60/30
P1.10	f	67	2	0	1	IV	AC	LUL	66/33
P1.11	m	63	2b	0	0	IIA	SCC	LLL	70/35
P1.12	m	61	3	2	0	IIIB	AC	LUL	60/30
P1.13	m	62	4	2	0	IIIB	SCC	LUL	66/33
P1.14	m	50	3	0	1	IV	SCC	RUL	66/33
P1.15	m	75	4	3	1	IV	SCC	RLL	50/25
		63.3±10.1 (49-88)							

For the lung and total body VOIs, the IVH of the sCT, CBCT, pCT, pCT-CBCT and pCT-sCT were generated and exported with velocity. The IVHs of the sCT and CBCT were compared with the one of the pCT. The IVH shows the frequency distribution of the gray values in the VOI. The more the IVHs are in alignment with the pCT, the more similar is their gray value distribution and the more likely can the CT-to-ED calibration be adopted by the pCT for that particular image. Each voxel belonging to the VOI of the sCT is compared to the corresponding voxel of the pCT by calculating the ME and MAE. The ME indicates the average error and the MAE is a measure of the dispersion of the error. The ME and MAE were calculated using the IVHs of the difference images with the following equations:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|CT_{analysis}(i) - CT_{reference}(i)\right| \tag{1}$$

$$ME = \frac{1}{N}\sum_{i=1}^{N}CT_{analysis}(i) - CT_{reference}(i) \tag{2}$$

N is the number of voxels in the VOI. $CT_{analysis}(i)$ and $CT_{reference}(i)$ are the i-th gray value of the analyzing and reference images, respectively [94]. Using the MAEs of the CBCTs and sCTs, the relative improvement of the MAE through the algorithm was calculated to compare the results with other studies.

Figure 2 - Volumes of Interest (VOI) for the intensity volume histogram generation. The total body VOI is illustrated in dark blue and the VOI of the lungs in light blue. To obtain the VOI for the lung and the VOI for the total body, respectively, a margin of 10 mm and 15 mm was subtracted from the lung and the patient outline, respectively, for possible anatomical differences between the two images.

2.2.3 Analysis of the dosimetric accuracy

An accurate dose calculation is highly important for treatment plan assessment in general and thus for ART as well, since the dosimetric benefit of ART should predominate the possible error. To assess how the CT number differences affect the dosimetric calculations on the sCTs an expert physician retrospectively delineated the patient's outline, ipsilateral and contralateral lungs, heart, esophagus, spinal cord, and the GTV on the pCT and sCT in a first step. For the PTV generation, 10 mm and 15 mm margins were added to the GTV in the axial and superior-inferior directions, respectively. Additionally, the new structure $lung_{ipsilateral}$-PTV was created by subtracting the PTV volume from the ipsilateral lung volume and referred to as $lung_{ipsilateral}$.

Initial VMAT treatment plans were created on the pCT with a grid spacing of 2 mm and a statistical uncertainty of 0.5 % using Monaco (Monaco 5.11, Elekta), a Monte Carlo based treatment planning system. The dosimetric constraints for treatment planning were limited by a dose a certain percentage x of a structure ($D_{x\%}$) or a certain volume y receives (V_{yGy}). The dosimetric constraints for the inverse treatment planning were: $V_{20Gy}(lung_{ipsilateral}) \leq 37\,\%$, $V_{20Gy}(lung_{total}) \leq 30\,\%$, $MLD \leq 20\,Gy$, $V_{5Gy}(heart)$ as low as possible, $V_{35Gy}(heart) < 10\,\%$, $MHD < 10\,Gy$, $D_{0.1\%}(spinal\ cord) < 50.5\,Gy$, and $MED < 34\,Gy$. The treatment plans were optimized for maximum dose compliance and then normalized so that the D_{Presc} covers 50 % of the PTV.

In the next step to evaluate dosimetric accuracy on the sCTs, the initial treatment plans from the pCTs were recalculated using identical beam settings on the corresponding sCTs. In cases were the FOV of the CBCT was too small and resulted in cropped images a structure for the missing part of the patient was generated by copying the external structure from the pCT to the CBCT and subtracting the FOV of the CBCT. The missing patient structure was copied to the sCT and the relative electron density to water (ED) of this structure was overwritten with 1 (Fig. 3) [95]. The dose distributions of the treatment plans on the sCTs were compared to the ones on the associated pCTs, with the pCTs as the reference images.
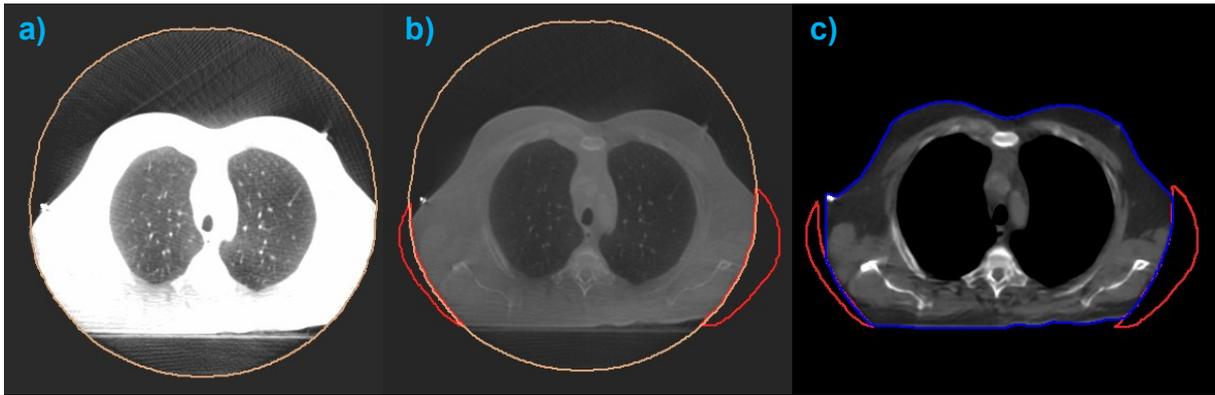
Figure 3 - Generation of the structure for the missing patient due to a too small field of view (FOV). a) First the FOV is contoured on the cone-beam CT (CBCT), b) then the structure of the patient outline from the CT is copied to the CBCT. A subtracted structure patient outline-FOV is created and referred to as the missing patient structure. C) The missing patient structure is copied to the synthetic CT and the electron density relative to water of the structure is overwritten with 1.

In this work, relevant dose volume histogram (DVH) parameters of the PTV (near-minimum ($D_{98\%}$), median ($D_{50\%}$) and near-maximum ($D_{2\%}$)) and of the OAR (V_{20Gy}(lung$_{ipsilateral}$), V_{30Gy}(lung$_{ipsilateral}$), V_{5Gy}(lung$_{contralateral}$), V_{4Gy}(heart), V_{25Gy}(heart), $D_{2\%}$(spinal cord), and MED as well as a global 2D and 3D gamma analysis with different gamma criteria are used to compare the dose distributions on the pCTs and sCTs. In addition to the gamma analysis, difference dose distributions pCT-sCT were created for each patient to allow better analysis where the dosimetric differences in dose distributions occur.

The DVH parameter analysis is highly dependent on the volume of the structures that may have changed in the period between the pCT and CBCT acquisitions. Thus the structures of the pCT were copied to the sCT_1 for the DVH analysis. Nevertheless, the changed density values do not exactly match the structures from the pCT due to the changed anatomy, which can still lead to discrepancies. In contrast the analysis of the percentage dose difference (DD) would be extremely sensitive to minor spatial displacements in regions with steep dose gradients, which can lead to misleadingly increased DD values [96]. The gamma analysis combines the distance with the dose criterion in one parameter and was therefore considered as more relevant than the DVH parameter analysis. The gamma value is still dependent on the patient shape, but it is less significant for the individual OARs.

The gamma analysis is a method developed by Low et al. specifically for the verification of IMRT treatment plans and performs a quantitative comparison between a reference dose distribution $D_{ref}(\overrightarrow{r_{ref}})$ and the dose distribution that should be evaluated $D_{eval}(\overrightarrow{r_{eval}})$ [96, 97]. This dose verification methodology includes two criteria, the DD and the distance-to-agreement (DTA). The DD describes the relative difference between the dose to be evaluated (D$_{eval}$) and the reference dose (D$_{ref}$). The DTA calculates the minimum distance to the next point of the reference dose distribution which has the same dose value for each pixel/voxel. Determined by the number of reference points available within the dose distribution under consideration, a large number of gamma values is obtained for each individual measurement point. It is possible to reduce this evaluation to a single value for each dose point, the gamma index. This is defined by the shortest distance from the dose distribution to be evaluated to the reference dose distribution, the minimum gamma value [96]. The smaller the gamma index is, the higher is the agreement of the two dose distributions for the corresponding measurement point. Furthermore, the agreement is within the defined tolerance if the gamma index is smaller than or equal to one. If, on the other hand, the gamma index is greater than one, a greater deviation than the defined tolerance for the measuring point under consideration exists.

The gamma index (γ), gamma value (Γ), DTA, and DD are described by the following formulas:

$$\gamma = min\{\Gamma\} \ \forall \ \{\overrightarrow{r_{eval}}\} \tag{3}$$

$$\Gamma = \sqrt{\frac{DTA^2}{\Delta d_{acc}^2} + \frac{DD^2}{\Delta D_{acc}^2}} \tag{4}$$

$$DTA = |\overrightarrow{r_{ref}} \text{-} \overrightarrow{r_{eval}}| \tag{5}$$

$$DD = D_{ref} - D_{eval} \tag{6}$$

with the reference and evaluating dose point position vectors $\overrightarrow{r_{ref}} \ and \ \overrightarrow{r_{eval}}$, the acceptable spatial deviation Δd_{acc}^2 and the acceptable dose deviation ΔD_{acc}^2.

24

In the context of plan verification, the percentage ratio of dose points fulfilling the gamma criterion to those that do not fulfill the gamma criterion, the gamma pass rate (GPR), is usually evaluated. The most commonly used GPR criteria are $\Delta d_{acc}^2/\Delta D_{acc}^2 = 3\,\%/3\,mm$ and $2\,\%/2\,mm$ in order to verify dose distributions or compare two treatment plans with each other. In this research study for plan comparison global 2D and 3D GPR were evaluated with respective gamma criteria of $2\,\%/2\,mm$ and $3\,\%/3\,mm$. Low dose thresholds, which define the smallest dose value that is included in the gamma analysis, were set to $25\,\%/40\,\%$ (2D-GPR) and $10\,\%/25\,\%$ (3D-GPR).

With a higher dose threshold, the focus is shifted to the high dose area, but this reduces the total number of voxels that are evaluated as seen in Fig. 4. It can also be seen that the number of voxels fulfilling the gamma criterion decreases as the stringency of the gamma criterion increases.

2.3 Image segmentation accuracy analysis on sCT

Another important part for ART is the structure generation and the time needed for it. The patient is already positioned for the treatment and any additional anatomical changes of the patient during the period between CBCT acquisition and irradiation would reduce the benefit of ART. To enable a fast image segmentation, the accuracy of a DIR on the sCTs was tested as a tool for ART. The generated structures were compared to the manually delineated structures by means of the dice similarity coefficient (DSC), Hausdorff distance (HD), mean surface distance (d_{mean}), sensitivity and specificity. The accuracy of the structures can on the one hand depend on the image quality and on the other hand on the chosen DIR algorithm. In addition, the time needed for the necessary structural adjustments after the DIR was recorded.

For the image segmentation analysis the same patient cohort as for the evaluation of the generated sCT, the image quality and dosimetric accuracy analysis, was used. The characteristics of the patients are displayed in Tab. 2.

Gamma criteria
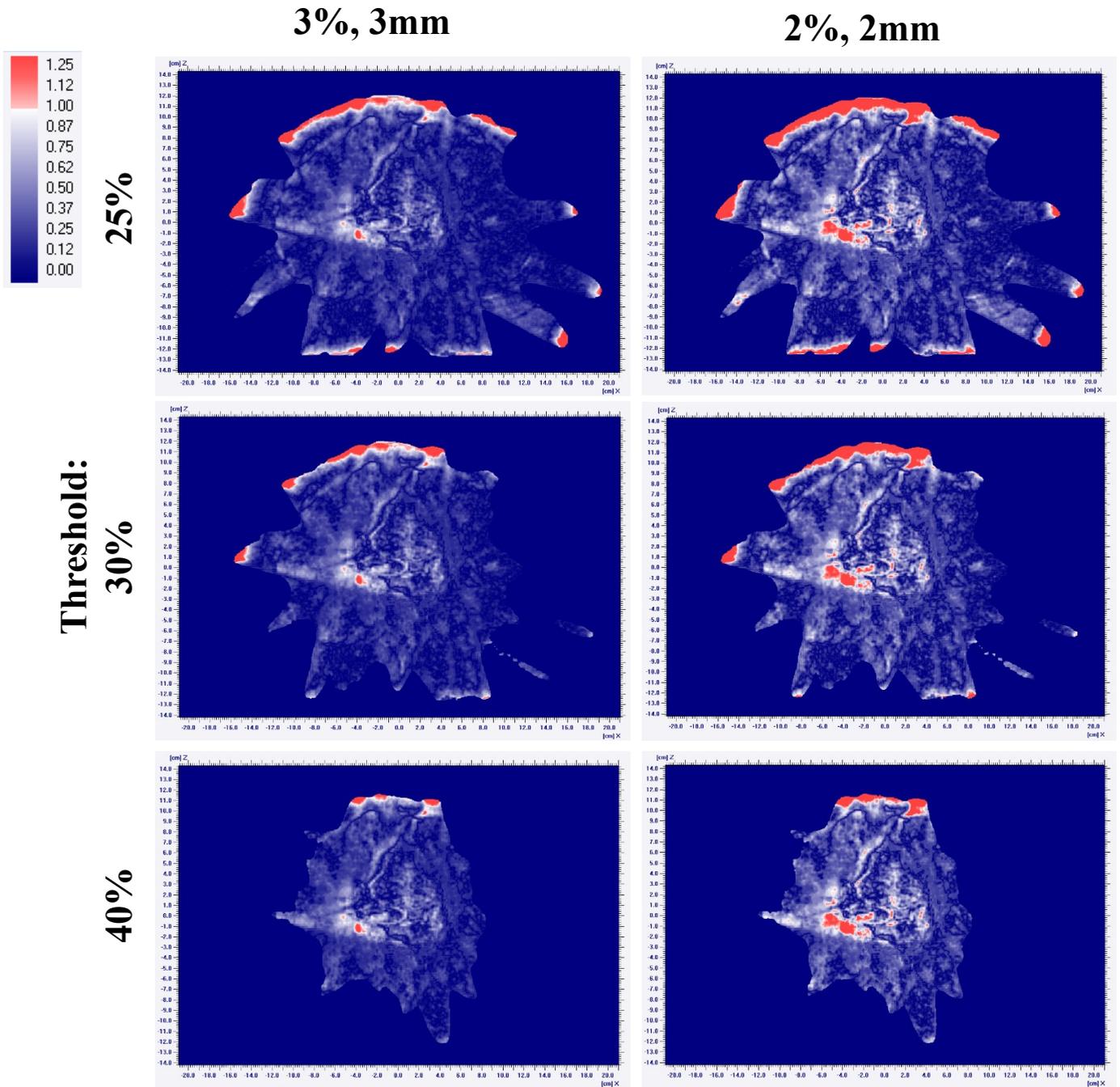
3%, 3mm

2%, 2mm

Figure 4 - Effect of the threshold of the gamma criterion. The threshold values 25 %, 30 %, and 40 % for the gamma criteria 2 %/2 mm and 3 %/3 mm are compared with each other. The smaller volume for analysis with increasing threshold can be seen. In addition, the number of voxels that do not fulfil the gamma criterion increases with the stricter gamma criterion of 2 %/2 mm.

26

The manually delineated OAR by the expert physician (patient outline, ipsilateral and contralateral lung, heart, esophagus, and spinal cord) of the pCT were deformable registered to the sCT using the intra-patient DIR module in ADMIRE AI. There are three registration steps in this DIR that gradually increase the freedom of image matching: Firstly a global rigid registration based on mutual information is implemented, secondly a block-wise matching with a normalized sum of squared differences is performed, and thirdly a dense DIR using local correlation coefficients to further approve the alignment is applied [98]. This results in a deformation vector for each voxel and a deformation vector field (DVF) for the entire image. Subsequently, the DVF is applied to the manually delineated structures on the pCT to generate new deformed structures on the sCT. Since the FOV of the CBCT and consequently also the FOV of the sCT is limited, all deformed structures were cropped in the craniocaudal direction if necessary. The DSC, HD, d_{mean}, sensitivity and specificity were used to compare the generated deformed structures (G) with the manually delineated ones (M).

The formulas for the five segmentation metrics are as follows:

$$\text{DSC} = \frac{2|M \cap G|}{|M| + |G|} \qquad (7)$$

$$\text{HD}(X, Y) = \max_{x \in X} \min_{y \in Y} \|x - y\| \qquad (8)$$

$$d_{mean} = \frac{\overrightarrow{d_{avg}}(X,Y) + \overrightarrow{d_{avg}}(Y,X)}{2} \qquad (9)$$

$$\text{Sensitivity} = \frac{|M \cap G|}{|M|} \qquad (10)$$

$$\text{Specificity} = \frac{|\overline{M} \cap \overline{G}|}{|\overline{M}|} \qquad (11)$$

With the voxels outside the structures M and G, $\overline{M}$ and $\overline{G}$.

The overlap of the manually delineated and deformed structures is measured by the DSC. The DSC ranges from 0 to 1, where a DSC of 0 means that the structures do not overlap at all, and a DSC of 1 represents two structures that are completely equal. The HD is a metric for the largest distance between two structures and is therefore susceptible to single large deviations. The mean surface distance determines how far, on average, a point of the structure X is away from its nearest neighbor in Y by averaging the two directed mean surface distances $\overrightarrow{d_{avg}}(X,Y)$ and $\overrightarrow{d_{avg}}(Y,X)$. Sensitivity and specificity indicate the percentage of a false positive or false negative structure. That means, the sensitivity shows how many percent of a structure really match the reference structure. The specificity, on the other hand, indicates the percentage by which the area outside the generated structure matches with that of the reference structure. It can be inferred from this that if the sensitivity is large and the specificity is small, the generated structure tends to be smaller than the reference structure and vice versa. Since each physician also delineates the structures slightly different, the results from the DSC and d_{mean} are also compared to the interobserver variabilities.

2.4 Dosimetric benefit of ART for patients with stage III NSCLC

Based on the previous analyses, the accuracy of dosimetric calculations on the generated sCTs was demonstrated. Therefore, in the next subproject, it is on the one hand possible to calculate the actually delivered dose distributions for IGRT and, on the other hand, to generate new treatment plans for the isoeffective and isotoxic ART scenarios. The actual delivered dose and the benefit of the 2 ART scenarios were retrospectively examined for a new patient population, which will be described in the next subchapter. A flowchart of the workflow for this part of the study can be seen in Fig. 5.
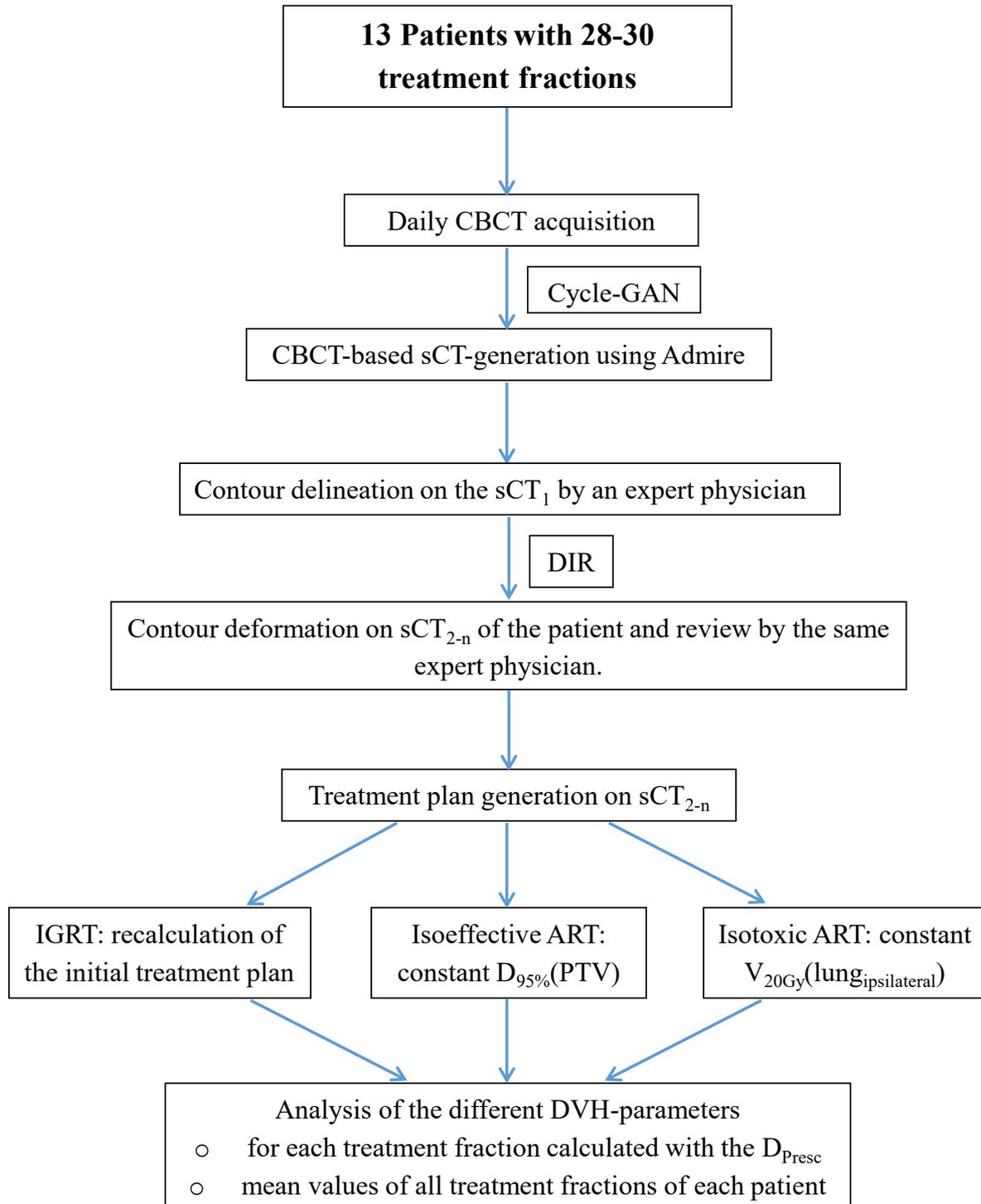
Figure 5 - Flowchart of the subproject analyzing the benefits of daily adaptive radiotherapy (ART). The individual steps from image generation to segmentation and dosimetric evaluation are shown. With CBCT=Cone-beam CT, cycle-GAN=cycle-generative adversarial network, sCT= synthetic CT, sCT_1=sCT of the first treatment fraction, DIR=deformable image registration, sCT_{2-n}= all treatment fractions except the first, IGRT=image guided radiotherapy, PTV= planning target volume, $D_{95\%}$(PTV)=the dose 95 % of the PTV receives, and V_{20Gy}(lung$_{ipsilateral}$)=the volume of the ipsilateral lung receiving 20 Gy.

2.4.1 Patient population

13 patients with stage III/IV NSCLC (staged according to AJCC TNM staging version 8) were included in this retrospective treatment planning study. Patient demographics are shown in Tab. 3. The study included primary tumors, but excluded secondary lesions of two patients. All patients underwent concurrent chemoradiotherapy (cCRT). The prescription dose was 56 Gy-60 Gy in 2 Gy fractions.

All patients were treated with a medical linear accelerator (VersaHD, Elekta AB, Stockholm, Sweden) with 10 MV volumetric modulated arc therapy (VMAT) treatment plans. A computer-controlled deep inspiration breath-hold (DIBH) technique was used for target immobilization during treatment planning CT acquisitions and each dose delivery. For 6 patients the Active Breathing Coordinator (ABC, Elekta AB, Sweden) and for 7 patients Catalyst (Catalyst, C-RAD, Sweden) was used for the DIBH. Directly before each patients' treatment daily kV-based CBCT scans (XVI 5.0, Elekta AB, Sweden) for the correction and verification of the patient positioning were obtained in several "stop-and-go" breath-hold phases. Because of the use of DIBH, no ITV margins were required, instead only GTV-PTV margins of 10 mm in the axial and 15 mm in the inferior-superior direction were used.

2.4.2 Study image data preparation

The imaging parameters for the pCT and CBCT generation were the same as those of the training datasets. All CBCTs of the patients were exported to the treatment planning system Monaco (Monaco 5.11, Elekta) with a slice thickness of 3 mm after reconstruction and the clinically used rigid registration to the pCT. These CBCTs were retrospectively converted into sCTs using ADMIRE AI. A DIR was also retrospectively applied on the manually delineated structures patient outline, ipsilateral and contralateral lung, heart, esophagus, spinal cord, and GTV from the sCT_1 to all remaining sCTs (sCT_{2-n}) for each patient using ADMIRE AI. The same expert physician, who delineated the structures on the pCT and sCT_1 retrospectively, reviewed all deformed structures and corrected them if necessary. For all sCTs of one patient, the same settings for window and level were used to review and correct the structures. The sCTs were not sorted by date when they were reviewed and corrected.

Table 3 - Patient population for the evaluation of the dosimetric benefit of daily offline ART for 13 patients with stage III non-small cell lung cancer. Listed are the patient's sex, age, tumor staging (staged according to AJCC TNM staging version 8), primary gross tumor volume (GTV) size on the first synthetic CT, carcinoma subtype, tumor location and the total prescription dose (D_{presc}), with f=female, m=male, SCC=squamous cell carcinoma, AD=adenocarcinoma, LLL=left lower lobe, RLL=right lower lobe, LML=left middle lobe, RML= right middle lobe, LUL=left upper lobe, RUL=right upper lobe, and fx= number of fractions. For the age and the primary tumor size additionally the average age and the range are listed.

Patient	Sex	Age	Tumor staging				Primary GTV (cm³)	Carcinoma Subtype	Tumor location	D_{presc} (Gy)/fx
			T	N	M	stage				
P2.1	m	70	4	0	1	IVA	81.7	SCC	RML	60/30
P2.2	m	60	4	3	0	IIIC	50.3	SCC	LLL	60/30
P2.3	m	83	3	3	0	IIIC	29.9	SCC	RUL	60/30
P2.4	m	59	3	1	0	IIIA	15.4	AC	RLL	60/30
P2.5	f	55	4	3	0	IIIC	163.8	AC	RUL	60/30
P2.6	m	80	4	3	0	IIIC	83.2	SCC	LUL	58/29
P2.7	m	80	4	0	0	IIIA	165.8	SCC	LUL	60/30
P2.8	m	73	3	3	0	IIIC	32.5	SCC	RUL	60/30
P2.9	f	69	4	2	0	IIIB	312.1	SCC	RUL	60/30
P2.10	m	50	4	2	1	IVA	82.4	SCC	RUL	60/30
P2.11	m	65	4	2	0	IIIB	231.2	SCC	LUL	60/30
P2.12	m	71	3	3	0	IIIC	116.6	AC	LML	56/28
P2.13	f	65	4	3	0	IIIC	83.9	AC	RUL	60/30
		67.7±9.6 (49-80)					111.4±83.0 (15.4-501.9)			

Like on the sCTs$_1$ the PTV and the structure lung$_{ipsilateral}$-PTV were created on all sCT$_{2-n}$ and will from now on be referred to as lung$_{ipsilateral}$. These margins were kept unchanged throughout the whole study and are conform to the study protocol of RTOG-0617.

The initial PTV on the first sCT (PTV$_1$) was transferred unaltered from the sCT$_1$ onto the sCTs$_{2-n}$ and moved to the center of the daily GTVs. This structure represents the initial MD that did not shrink synchronously with the GTV but changed the position according to the GTV movement.

In order to analyze the tumor regress or progress over the treatment period, the volume of the GTV was mapped against the fraction number for every patient.

2.4.3 Treatment planning and dosimetric analysis

Treatment plans were generated using a Monte Carlo based treatment planning system (Monaco 5.11, Elekta) with a 3 mm grid spacing and 1 % statistical uncertainty per dose calculation. The dosimetric constraints are listed in Tab. 4 and are adopted from the RTOG-0617 study protocol [9] except for the heart. Due to the potential association with cardiac events of the V$_{5Gy}$(heart), V$_{35Gy}$(heart), and MHD [35, 99, 100], the heart constraints were further reduced in comparison to RTOG-0617. A general initial template was created for the first treatment plan, which was only slightly modified for each patient. The IMRT constraints of the template are shown in Tab. 5. The treatment plans were normalized, such that the D$_{Presc}$ covers 95 % of the PTV. And were all clinically acceptable.

For the IGRT scenario, the initial treatment plan was recalculated for each patient onto their sCT$_{2-n}$ with identical beam and control point settings, grid size, statistical uncertainty, and the isocenter resulting from the image registration. This scenario illustrates the daily delivered dose of the patient with IGRT when neglecting intrafractional motion in DIBH and considering only interfractional anatomical changes. To compare this to the isotoxic and isoeffective ART approaches, two new treatment plans were generated by reoptimization on sCT$_{2-n}$. For the reoptimization, new isocenters were placed in the center-of-mass of the daily adjusted PTVs$_{2-n}$.

Table 4 – IMRT constraints for the treatment planning. The constraints are adopted from the RTOG-0617 study protocol except for the heart. Due to the potential association with cardiac events of the V_{5Gy}(heart), V_{35Gy}(heart), and mean heart dose (MHD), the heart constraints were further reduced in comparison to the RTOG-0617. With PTV=planning target volume, MED=mean esophagus dose, and MLD=mean lung dose.

Structure	Constraints		challenging cases
PTV	D_{min}	≤ 95 %	≤ 93 %
	D_{max}	≤ 72 Gy	≤ 75 Gy
	V_{60Gy}	≥ 95 %	≥ 90 %
Spinal cord	$D_{0.1\%}$	≤ 50.5 Gy	
Lung$_{total}$	V_{20Gy}	≤ 30 %	
Lung$_{ipsilateral}$	MLD	≤ 20 Gy	
	V_{20Gy}	≤ 37 %	
Esophagus	MED	< 34 Gy	
Heart	V_{60Gy}	< 33.3 %	
	V_{45Gy}	< 66.7 %	
	V_{40Gy}	< 100 %	
	MHD	< 10 Gy	
	V_{30Gy}	< 10 %	

Table 5 – Isoconstraints of the initial template, which was only slightly modified for each patient for the first treatment plan. With GTV=Gross tumor volume, PTV=Planning target volume, QU=Quadratic underdose, QO=Quadratic overdose, and EUD=Equivalent uniform dose.

Structure	Cost function	Reference Dose (cGy)	Power law exponent	shrink margin (cm)	Isoconstraint
GTV	QU	6000			10
PTV	Target EUD	6000			
	QO	6300			100
	QU	6000			50
Lung$_{ipsilateral}$	Serial		2	0.3	2200
	QO	3500		0.3	100
Lung$_{contralateral}$	Serial		2	0.3	1200
Heart	Serial		2	0.3	1300
Spinal cord	QO	2000		0.3	80
Esophagus	QO	2500		0.3	100
Patient	QO	3000		1.5	110

For the isoeffective approach, the ipsilateral lung dose constraints of the initial prescription template were modified by prescribing progressively more stringent for as long as the PTV remained adequately covered. Each treatment plan was normalized just as in the initial plan, so that the D_{Presc} covers 95 % of the daily PTV_{2-n}, for the isoeffective scenario.

For the isotoxic scenario, an equivalent process was used with an increase of the prescription dose for the PTV_{2-n}. The $V_{20Gy}(lung_{ipsilateral})$ should not exceed the one of the initial treatment plan using this procedure. For this reason, the treatment plans were normalized to the V_{20Gy} of the $lung_{ipsilateral}$ of the initial treatment plan on sCT_1 for the isotoxic approach. The normalization ranged from 90 % to 110 %. In cases were the ipsilateral lung was cropped due to the limited FOV of the CBCT (Fig. 6), the volume of the initial ipsilateral lung was used for the normalization of the isotoxic treatment plans. This was the case for three patients. Furthermore, a maximal D_{Presc} for each treatment fraction of 3.3 Gy was established, in agreement with the 2.2 Gy-3.8 Gy of the RTOG-1106 study protocol, as their study showed no adverse effects for this dose escalation [101]. In order to avoid uncertainties caused by a DIR-based dose accumulation, the DVH parameters of each fraction were evaluated separately. To be able to use the standard DVH parameters for the evaluation of the cumulative dose distribution (such as V_{20Gy}), each fraction was optimized and evaluated with the total D_{presc}. Therefore, treatment plans of the isotoxic ART scenario, in which a prescription dose to the PTV of more than 100 Gy was achievable, were normalized to cover 95 % of the PTV with 100 Gy.

Dose distributions of all treatment fractions were analyzed in terms of $MD_{initial}$, GTV and PTV coverage ($D_{95\%}$), V_{20Gy} of the ipsilateral lung and MLD of the ipsilateral and contralateral lung, MHD and the V_{5Gy} and V_{30Gy} of the heart, MED, and $D_{0.1\%}$ of the spinal cord. Furthermore, the linear quadratic model was used to calculate the equivalent dose in 2 Gy-fractions (EQD_2) for the $D_{95\%}(MD_{initial})$, $D_{95\%}(GTV)$ and $D_{95\%}(PTV)$ using a specific tissue characterization ratio α/β of 8.2 Gy [102]. The aforementioned dosimetric parameters of the initial plans were compared to the resulting plans of the IGRT, isoeffective, and isotoxic scenario. Additionally, the number of treatment fractions violating the constraints were analyzed.
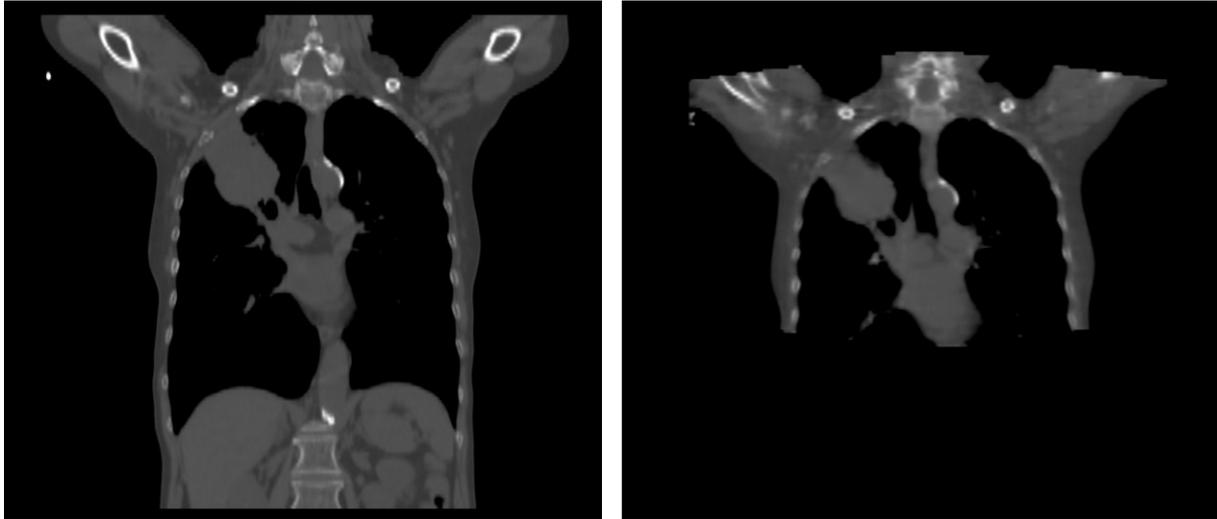
Figure 6 – Treatment planning CT (left) and synthetic CT (right) of a patient whose lungs were cropped on the cone-beam CT and, therefore, also in the synthetic CT.

The dose accumulation was estimated using the mean of all treatment fractions for each patient to analyze the number of patients violating the dosimetric constraints. The $D_{95\%}$(GTV) and $D_{95\%}$(PTV) were considered as violated if the $D_{95\%}$ was less than 95 % of the D_{Presc}. The $D_{95\%}$($MD_{initial}$), however, was considered as violated when less than 95% was covered with 80 % of the D_{Presc} because the MD area has a lower density of tumor cells and needs less dose [103]. Furthermore, we used the initial PTV for the calculation of the MD and not the smaller clinical target volume (CTV), in which the entire MD should already be included.

2.4.4 Statistical analysis

All results were visually tested for normal distribution using a quantile-quantile plot (Q-Q plot). In the Q-Q plots, the quantiles of the measured values are compared to the theoretical quantiles that the data should have if they were perfectly normally distributed. In the case of a perfect normal distribution, all expected values would be located on the straight line. The more they deviate from the line, the less the data is normally distributed. Additionally, if all or almost all of the data is within the 95 % confidence interval, this is an indicator that the data is normally distributed. In addition, the Kolmogorov-Smirnov test was performed with a significance level of 0.05. If the p-value is smaller than 0.05, the hypothesis of a normal distribution is rejected. For normally distributed results, mean±SD was given; for non-normally distributed, mean±interquartile range (IQR) was reported instead.

In the image uncertainty analysis the mean ME values were, in addition, compared with 0 HU using a two-tailed t-test to differentiate if the noise distribution is Gaussian distributed around 0 HU or if it is a real noise distribution. Statistical significance was demonstrated for p-values lower than 0.05.

For the image segmentation analysis the results of the ipsilateral and contralateral lung were tested for statistically significant differences using a two-tailed t-test. Statistical significant differences were found for p-values below 0.05.

For the dosimetric accuracy analysis, all previously mentioned DVH parameters of the IGRT, isoeffective ART, and isotoxic ART treatment plans were compared with the initial treatment plan and analyzed for statistically significant dosimetric differences using a nonparametric Wilcoxon signed-rank test in R (RStudio 1.4.1717, PBC, Boston, MA). Statistical significance was achieved for p-values < 0.05.

Furthermore, Pearson correlation coefficients (PCC) were calculated to analyze possible correlations between GTV regression and the analyzed DVH parameters. The correlation coefficient ranges between -1 and 1. Positive values are referred to as a positive correlation. This indicates that the analyzed variables have a positive linear correlation to each other and thus an increase of one variable corresponds to an increase of the other variable. Negative values, on the other hand, indicate a negative correlation. In this case, an increase of one variable causes a decrease of the other. The closer the value is to +1 or -1, the stronger is the degree of positive or negative linear correlation. If the value is close to 0, there is no correlation between the two variables.

3. Results

The results are divided into subchapters about the evaluation of the image quality of the generated sCTs (image uncertainties and dosimetric accuracy), image segmentation accuracy, tumor size variation and finally the dosimetric benefit of ART for NSCLC patients (with isoeffective ART and isotoxic ART compared to the initial treatment plan and to the scenario without ART).

3.1 Evaluation of the generated thoracic sCTs' image quality

An exact assignment of the grey values to the relative electron density is important to ensure a precise dose calculation on the sCTs. Therefore, before the generated sCTs can be used for dosimetric calculations, their image quality must be evaluated. For this purpose, the $sCTs_1$ were compared to the pCT regarding the image uncertainties and dosimetric accuracy.

3.1.1 Image uncertainties analysis of the generated sCT

Difference images and IVH of the generated sCTs compared to the CBCTs and pCT

Fig. 7 shows an axial (left column), coronal (middle column), and sagittal slice (right column) of a representative patient of the pCT (first row), the corresponding unprocessed $CBCT_1$ (second row), and the resulting sCT_1 (third row). For all images the same window and level (window: 1000 HU, level: -100 HU) were used. When comparing the generated sCT with the CBCT on a visual basis one can see that the shapes of the CBCTs are preserved in the sCT and have not been distorted. The gray values and contrast of the sCT match the pCT more than the CBCT especially in the area of the ribs, heart, and also in the adipose tissue, and muscle and lean tissue. The structures are better distinguishable in the sCT, but still show slight smearing in the sagittal view compared to the pCT. In addition, fewer streak like artifacts are seen in the sCT.
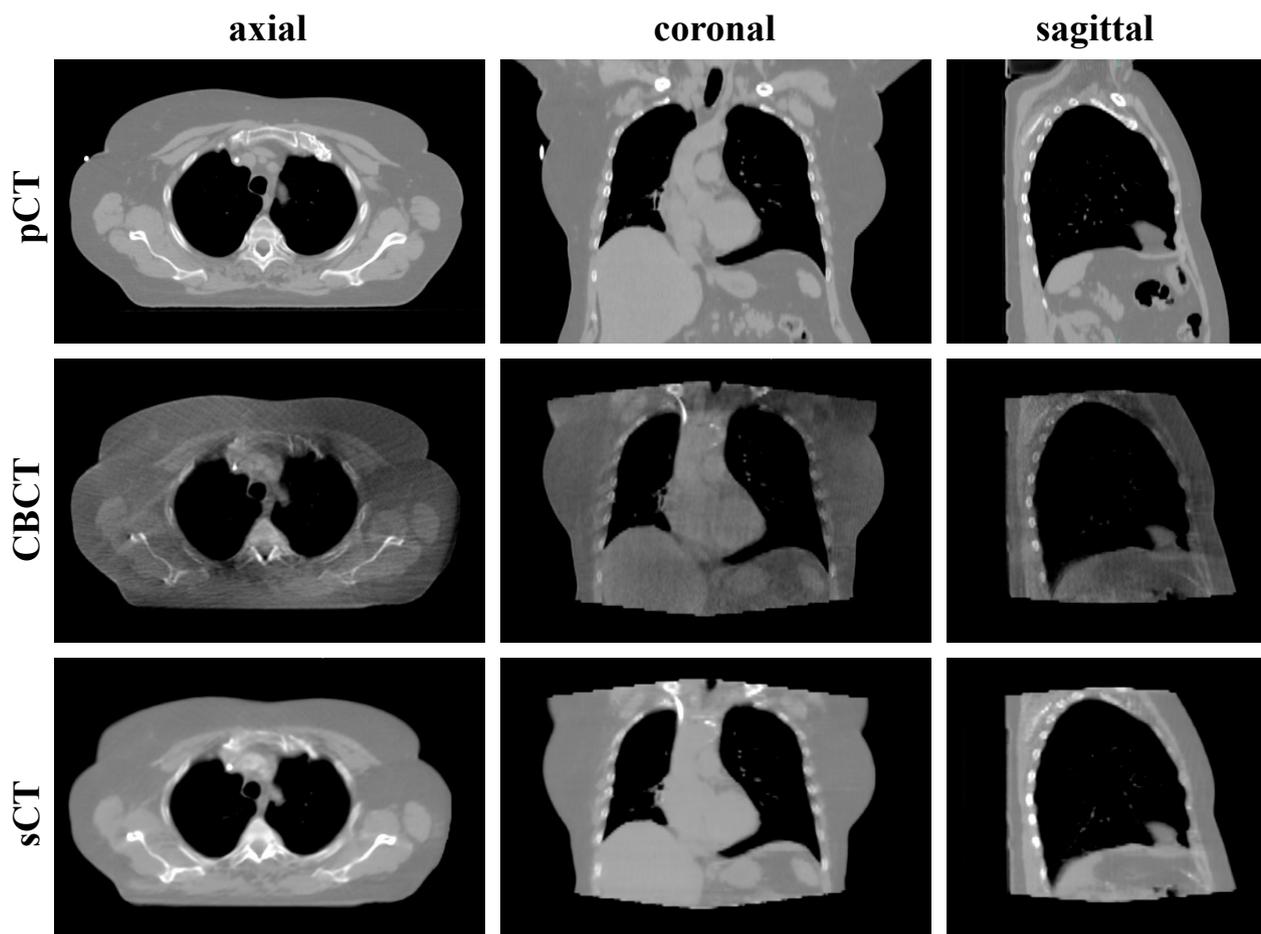
Figure 7 - The planning CT (pCT), a cone-beam CT (CBCT) and a generated synthetic CT (sCT) in comparison for the axial, coronal and sagittal view for a representative patient. The same window and level settings were used for the images (window: 1000, level: -100).

In the difference images shown in Fig. 8, darker gray levels than the background indicate too low and lighter ones too high CT numbers. The white and black parts - for example at the patient and lung outline - result from anatomical changes possibly caused by a slightly different patient positioning or weight gain or loss of the patient. These differences can occur due to the different acquisition days of the images.

Nevertheless, in the difference image CBCT-pCT one can see, that the grey levels inside the lungs are lighter and, therefore, the CT numbers inside the lungs are clearly higher in the CBCT than in the pCT. The CT numbers of the adipose tissue and muscle and lean tissue are in contrast lower in the CBCT compared to the pCT for both. This was noticeably improved in the sCT, so that the contrast in the differential image sCT-pCT is lower and most of the darker and lighter parts seem to occur due to different patient weight or positioning.
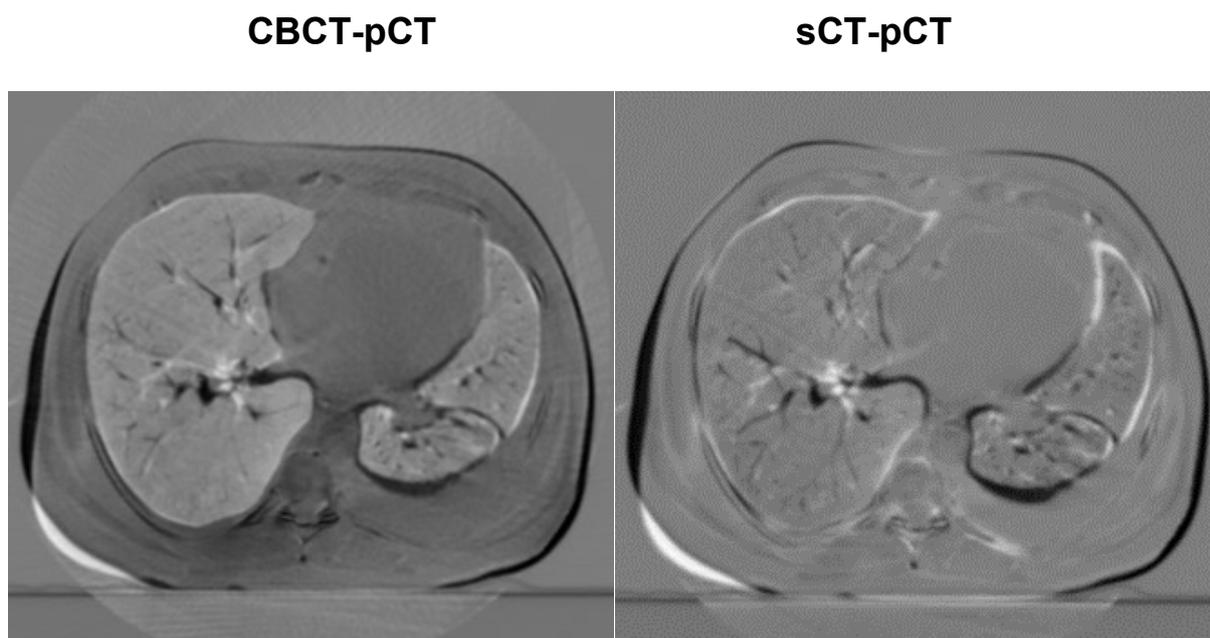
CBCT-pCT   **sCT-pCT**

Figure 8 - Difference images. The planning CT (pCT) was subtracted from the cone-beam CT (CBCT) (left) and the synthetic CT (sCT) (right), respectively. Darker gray levels than the background show that the CT number is too low and lighter ones that the CT number is too high. The white and black areas, are due to anatomical changes that might have been caused by a slight change in patient positioning or patient weight gain or loss.

This effect is also evident in the IVH. Fig. 9 shows an IVH of a representative patient (P1.3 – Fig. 9 a)), the worst (P1.7 – Fig. 9 b)) and the best (P1.11 – Fig. 9 c)) matching IVH of the sCT to the pCT. In this figure, the pCT is shown in blue, the sCT in green and the CBCT in red. In all three IVHs of the pCTs and sCTs, 3 peaks between -800 HU and -900 HU, -100 HU and -150 HU, and between 0 HU and 50 HU are evident and represent the voxels in the lung [44], adipose tissue, and muscle and lean tissue [104]. The IVHs of the CBCTs are more shifted than the IVHs of the sCTs, the peaks are broader and less high and the peak of the muscle and lean tissue cannot be separated from those of the adipose tissue. At the most, one could interpret a shoulder/plateau shifted by approximately 50 HU-100 HU to the second maximum as a third maximum in 9 patients. The shifts of the IVH-peaks were all normally distributed with p-values of the Kolmogorov-Smirnov test for the sCT of 1.00, 0.39, and 1.00 and for the CBCT of 0.60, 1.00, 0.97, for the lung, adipose tissue, and muscle and lean tissue peak, respectively.
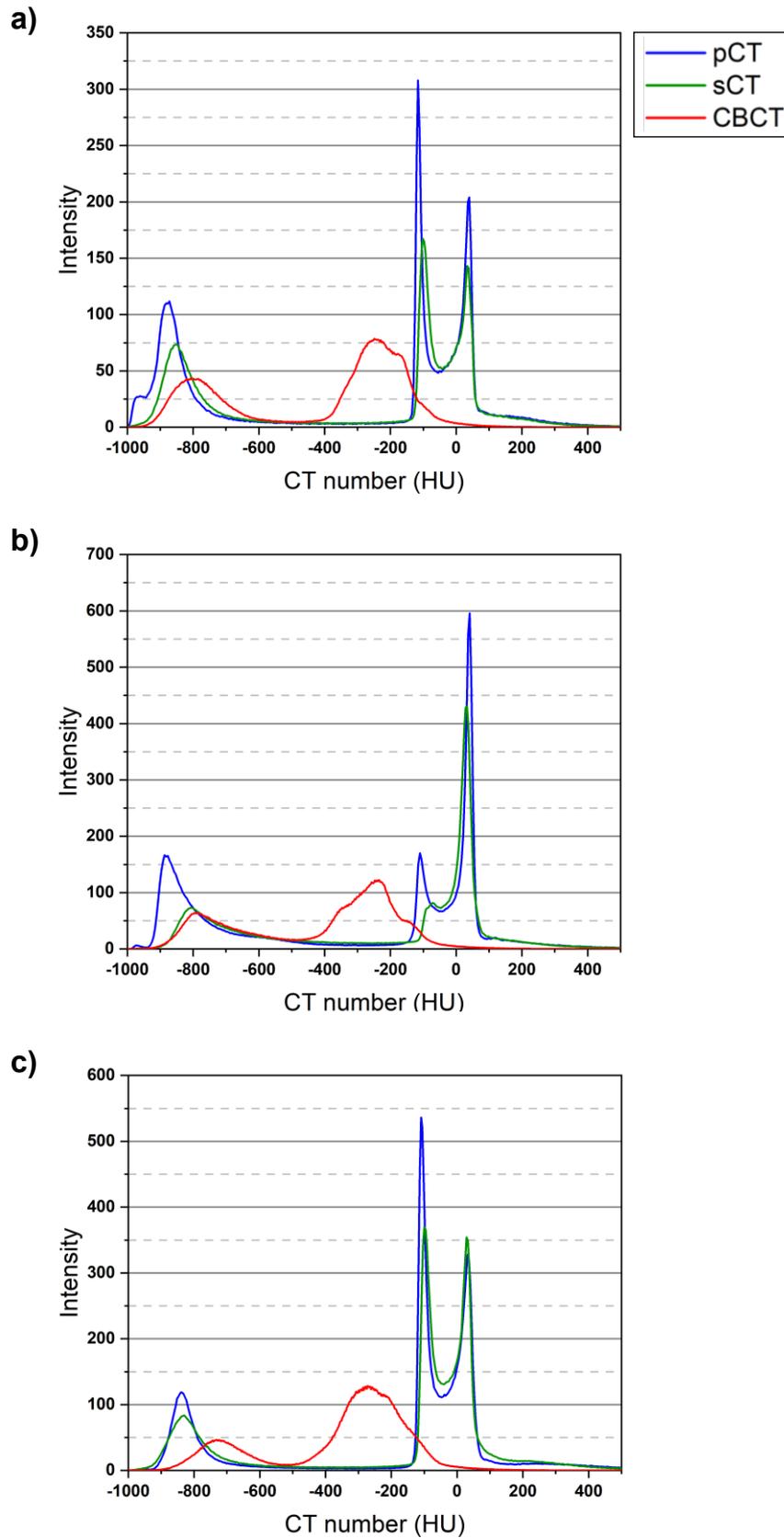
Figure 9 - Intensity-volume histograms (IVHs) a) of a representative patient (P1.3), b) the IVHs of synthetic CTs (sCTs) with the worst (P1.7) and c) best match (P1.11) to the IVHs of the planning CTs (pCTs) compared to the respective IVHs of the cone beam CTs (CBCTs).

The average deviation of the peaks of all sCT IVHs to pCT IVHs was 29.0 HU±24.7 HU, 19.2 HU±9.9 HU and -4.3 HU±8.1 HU with values in the ranges between -7 HU-85 HU, 9 HU-41 HU and -22 HU-10 HU for the lung, adipose tissue, and muscle and lean tissue, respectively. The IVHs of the uncorrected CBCTs had mean deviations of 73.9 HU±48.1 HU ranging between -3.0 HU and 167.0 HU, -119.1 HU±43.1 HU in the range between -200.0 HU and -6.0 HU and -174.5 HU±139.1 HU varying between -268.0 HU and -70.0 HU)) for the lung, adipose tissue, and muscle and lean tissue, respectively. Regarding the peak of the muscle and lean tissue, it must be taken into account that it was only visible in 9 patients and no third maximum was seen at all in the other 6 patients. That explains the high standard deviation of this peak. Except for one patient all peaks of the adipose tissue and all peaks of the lung are shifted in the positive direction in the IVHs of the sCTs. Whereas in 10 of the 15 patients, the peaks of muscle and lean tissue were shifted in the negative direction in the IVHs of the sCTs. For the IVHs of the CBCTs, it was the same with the lung peak. One was shifted in the negative direction, all others in the positive direction. For the adipose tissue peak and muscle and lean tissue peak, however, all peaks were shifted in the negative direction in the IVHs of the CBCTs.

The peaks in the IVH of the sCT of the representative patient (Fig. 9 a)) were shifted by 34 HU (lung peak), 15 HU (adipose tissue peak), and -6 HU (muscle and lean tissue peak) with respect to the IVH of the pCT. The peaks were all smaller than those of the pCT, but higher compared to the ones in the IVH of the CBCT. The shift of the peaks from the IVH of the CBCTs was also much more evident than for the IVH of the sCT with 80 HU, -133 HU, and -268 HU for the lung, adipose tissue, and muscle and lean tissue, respectively.

Fig. 9 b) shows the worst matching IVHs of the sCT and pCT. The peaks of the lungs, adipose tissue, and muscle and lean tissue are shifted by 71 HU, 41 HU and -8 HU in the sCT IVH, respectively. Nevertheless, the IVH of the sCT is much closer to that of the pCT than that of the CBCT, which has deviations from the location of the peaks of 80 HU (lung peak), -124 HU (adipose tissue peak), and -201 HU (shoulder of the muscle and lean tissue).

The location of the peaks of the best fitting sCT IVH (Fig. 9 c)), differ less than 10 HU, with only the intensity of the peaks being lower at the lung and adipose tissue peaks compared to the pCT IVH. In the IVH from the corresponding CBCT, the peaks are broader, the positions are clearly shifted compared to the pCT and differ by 107 HU (lung peak) and -163 HU (adipose tissue peak). Considering the shoulder as the third peak of the muscle and lean tissue, there is a deviation of -181 HU.

Mean error and mean absolute error of the generated sCTs compared to the pCTs

All results were normally distributed with p-values of the Kolmogorov-Smirnov test ranging between 0.26 and 0.91. The observation from the previous chapter is confirmed by the results for the ME and MAE of the two VOIs. The mean ME and MAE and the individual MEs and MAEs for each patient can be seen in the box plots and bar graphs in Fig. 10. The MAEs of the total body VOIs were improved by sCT conversion for each patient, whereas for the lung VOI, three patients had a slightly higher MAE compared to the CBCT. The MEs of the total body VOIs of the CBCT are for 14 of the 15 patients in the negative range and for the lung VOI for 11 of the 15 patients in the positive range, supporting the previous finding based on the difference images and the IVHs. The VOIs of the unprocessed CBCTs had a mean ME for the lung and total body VOIs of 50.1 HU±49.0 HU and -102.7 HU±53.6 HU and an MAE of 91.3 HU±27.4 HU and 190.2 HU±35.2 HU, respectively. After the sCT conversion, the mean MEs and MAEs for the VOIs were 22.3 HU±27.7 HU (ME for the lung VOI), 29.6 HU±28.9 HU (ME for the total body VOI), 73.1 HU±18.7 HU (MAE for the lung VOI) and 93.6 HU±29.0 HU (MAE for the total body VOI). This results in a MAE improvement of 19.9 %±32.4 % (lung VOI) and 50.5 %±14.7 % (total body VOI). The mean MEs were statistically significantly different from 0 HU. The SD of the MEs and MAEs was lower for both VOIs after sCT conversion compared to those of the unprocessed CBCTs. The SD of the pCT reference image noise was 11.4 HU.
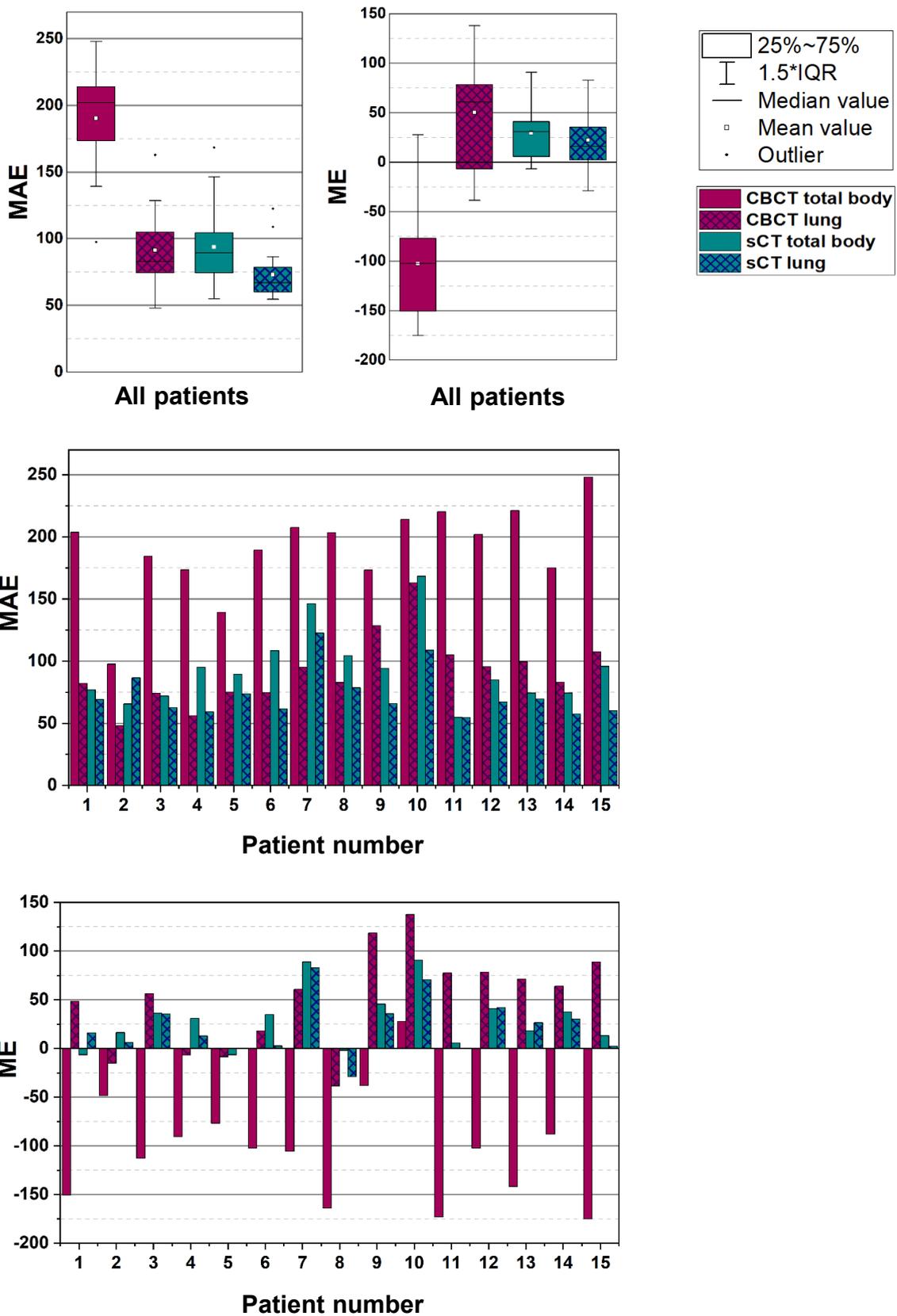
Figure 10 - Boxplots of the mean absolute error (MAE) and mean error (ME) for all patients (at the top) and bar graphs for all individual patients (in the bottom).

3.1.2 Dosimetric accuracy analysis of the generated sCTs

DVH parameters

The results were all normally distributed with p-values of the Kolmogorov-Smirnov test of 0.47 ($D2\%(PTV)$), 0.60 ($D50\%(PTV)$), and 1.00 ($D98\%(PTV)$). Fig. 11 shows an example of a Q-Q plot of the $D_{50\%}(PTV)$. All data points are inside the confidence interval, which indicates normally distributed data. The DVH parameters results are displayed in Fig. 12. Overall, the $D_{2\%}(PTV)$, $D_{50\%}(PTV)$, and $D_{98\%}(PTV)$ of the sCTs are all close to those on the pCT. The relative dose differences are within an absolute interquartile range of $\pm2.2\%$. The relative mean dosimetric deviations from the treatment plans on the sCT to the treatment plans on the pCT were $1.7\%\pm1.4\%$, $0.8\%\pm0.8\%$, and $0.9\%\pm1.2\%$ for the $D_{2\%}(PTV)$, $D_{50\%}(PTV)$ and $D_{98\%}(PTV)$, respectively. The relative DVH parameters of the OAR of the sCT differ by $-0.1\%\pm1.6\%$ ($V_{20Gy}(lung_{ipsilateral})$), $1.1\%\pm2.2\%$ ($V_{30Gy}(lung_{ipsilateral})$), $0.0\%\pm1.3\%$ ($V_{5Gy}(lung_{contralateral})$), $-0.1\%\pm1.4\%$ ($V_{4Gy}(heart)$), $0.4\%\pm1.3\%$ ($V_{25Gy}(heart)$), $0.2\%\pm1.0\%$ ($D_{2\%}(spinal\ cord)$), and $0.3\%\pm1.4\%$ (MED), from those on the pCT.

In 12 of the 15 patients all evaluated DVH parameters of the PTV were higher in the treatment plans of the sCTs than in the initial treatment plan. The OAR doses, on the other hand, were lower than in the initial treatment plan in 47.3% of the evaluated structures. Overall, it must be taken into account that the dose distributions are compared on images with possibly different anatomy and thus different density values. The altered anatomy has a major impact on the DVH parameters and thus the dosimetric deviation may be overestimated as a result. Furthermore, the cropped lungs also have an influence on the DVH parameters. There was an outlier for the $D_{2\%}$ of the PTV with a deviation of 6.1%. The deviation in the $D_{2\%}(PTV)$ could indicate small hotspots, which are less noticeable in the $D_{50\%}(PTV)$ and $D_{98\%}(PTV)$ because of their small size compared to the size of the PTV. In the case of the lungs and the heart, it is noticeable that the $V_{30Gy}(lung_{ipsilateral})$ and $V_{25Gy}(heart)$ each show a larger mean deviation and have larger whiskers than the $V_{20Gy}(lung_{ipsilateral})$ and $V_{4Gy}(heart)$. This is probably due to the fact that the absolute values are lower in each case and thus even small deviations account for larger percentage errors. The absolute deviation for these DVH parameters were $0.2\%\pm0.5\%$ ($V_{30Gy}(lung_{ipsilateral})$) and $0.0\%\pm0.2\%$ ($V_{25Gy}(heart)$).

45

Among the OAR, the ipsilateral lungs have the largest error bars, probably due to anatomical volume variations. For V_{4Gy}(heart), there was an outlier of about 4 %. This outlier occurred in the same patient as the outlier of the $D_{2\%}$(PTV). This patient had a significant weight loss between the pCT and CBCT image acquisitions. Furthermore, the tumor was shifted towards the heart and the heart volume was smaller by 16.4 %. This might be caused by a radiation-induced [105] or chemotherapy-induced [106] heart disease, but also by contouring inaccuracies due to the poor contrast in the periphery of the CBCT scan where the heart is located.
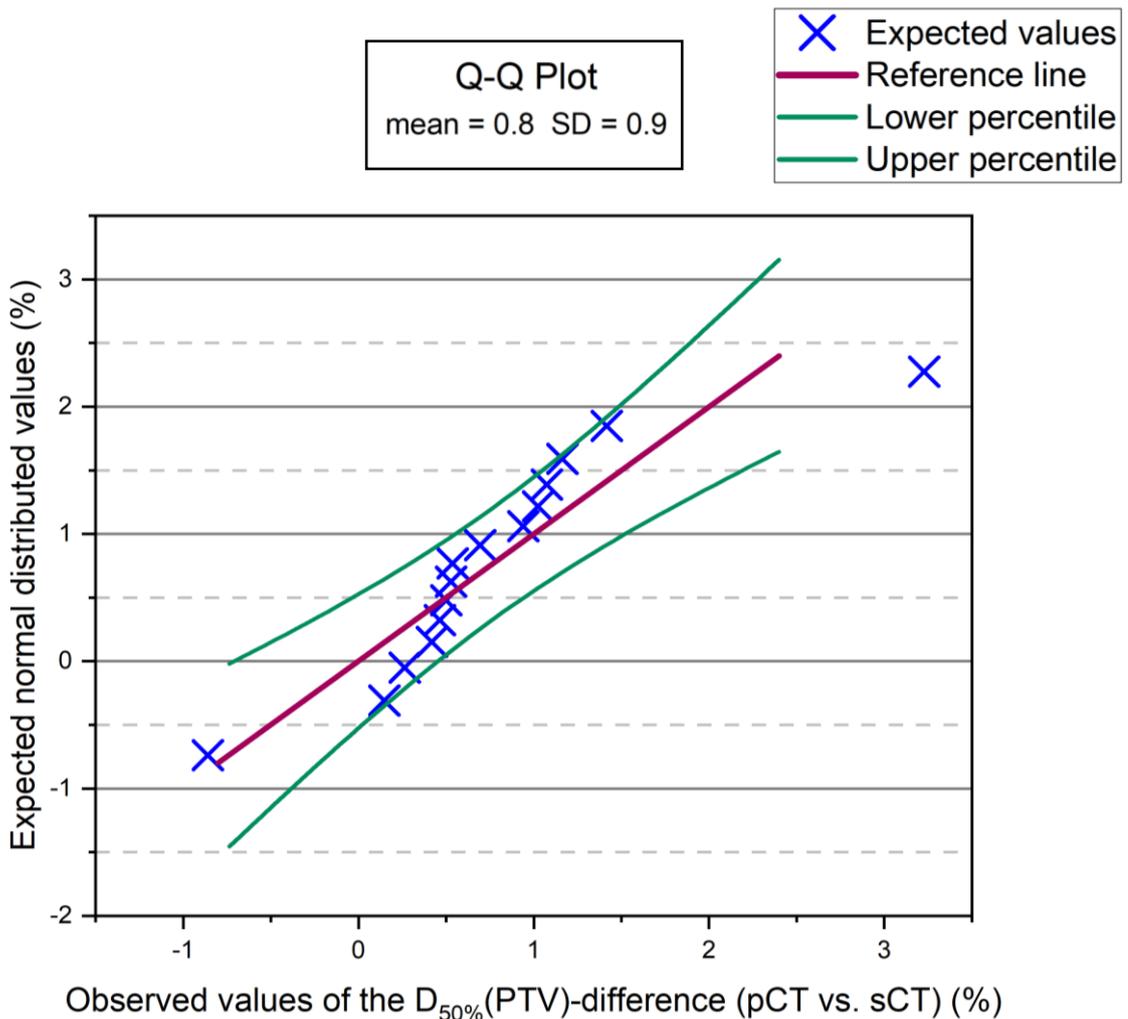
Figure 11 - Quantile-Quantile plot (Q-Q Plot) to test if the results of the difference of the $D_{50\%}$(PTV) between the planning CT (pCT) and synthetic CT (sCT) are normally distributed. In the Q-Q plots, the quantiles of the measured values are compared to the theoretical quantiles that the perfectly normally distributed data would have. In the case of a perfect normal distribution, all expected values would be on the reference line. The more they deviate from the reference line, the less the data is normally distributed. Additionally, if all or almost all of the data is within 95 % confidence interval (lower and upper percentile), as seen for the results, this is an indicator that the data is normally distributed. With SD=standard deviation.
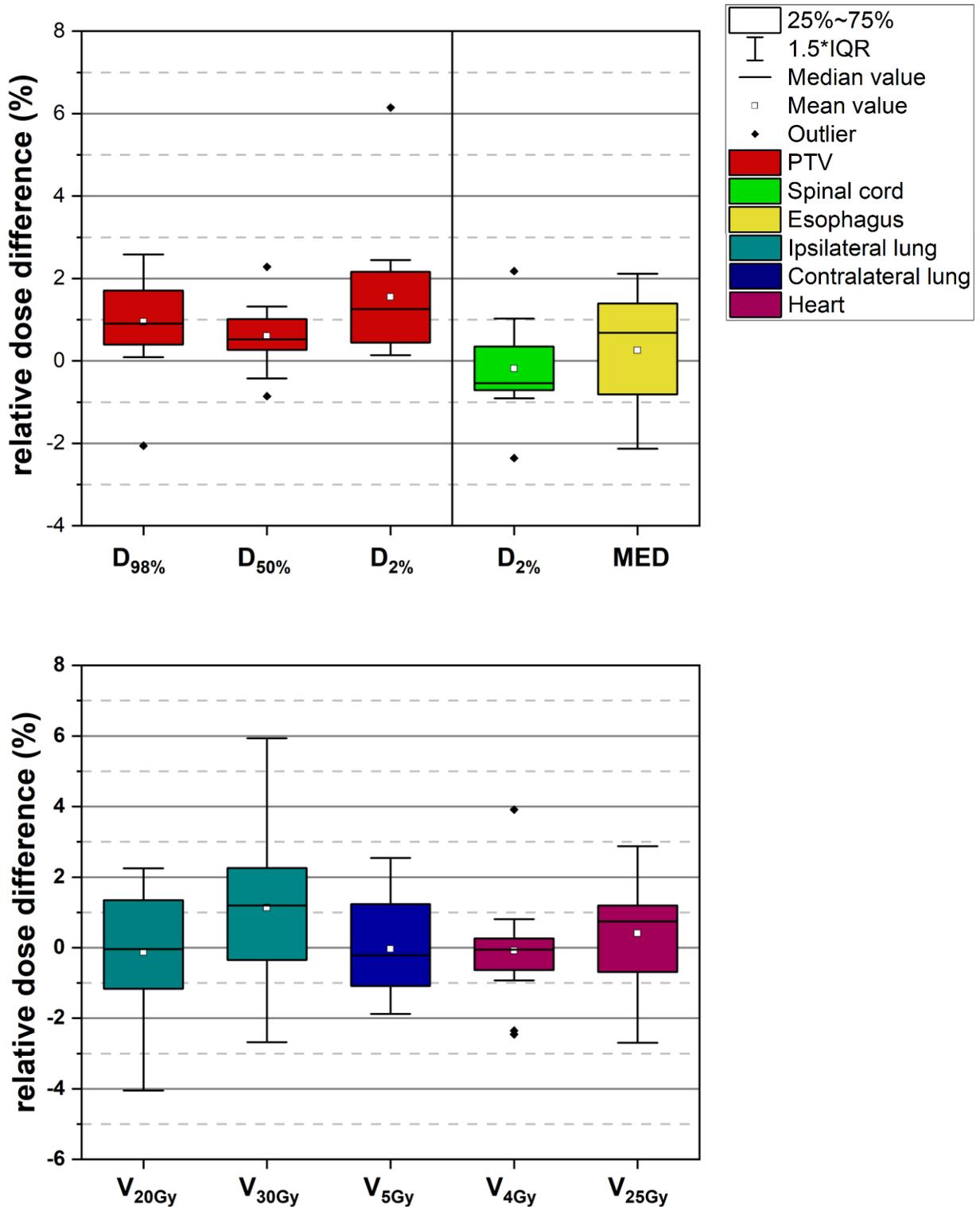
Figure 12 - DVH parameter analysis for the planning target volume (PTV) and organs at risk. Shown are the deviations of the $D_{2\%}$, $D_{50\%}$ and $D_{98\%}$ of the PTV, the $D_{2\%}$(spinal cord), mean esophagus dose (MED), V_{20Gy}(lung$_{ipsilateral}$), V_{30Gy}(lung$_{ipsilateral}$), V_{5Gy}(lung$_{contralateral}$), V_{4Gy}(heart), and V_{25Gy}(heart) compared to the DVH parameters of the initial treatment plan, respectively. Error bars include the data points within the interquartile range times 1.5 based on our patient population of 15 patients.

Global gamma pass rates

The global 2D and 3D GPR were not all normally distributed as seen in Fig. 13. The p-values of the Kolmogorov-Smirnov test ranged between 0.02 and 0.23. Therefore, the median values±IQR must be examined in this case. The global GPR of the different criteria were above 98.8%±1.6% (2D, 3%/3mm), 96.7%±2.8% (2D, 2%/2mm), 97.0%±1.8% (3D, 3%/3mm), and 94.6%±3.3% (3D, 2%/2mm). The GPR of each patient as well as the median±IQR are shown in Tab. 6. In all cases except for the global 2D gamma (2%/2mm) with the 25% threshold, the median value is better than the mean, showing that there are some outliers with worse GPR. The different thresholds indicate whether the dosimatic errors are located mainly in the low dose or in the high dose region. Since the median values of the GPR for each criterion improve with a higher threshold, the voxels with too large deviations not meeting the gamma criterion seem to be more in the low dose range at the periphery of the patient. This might be caused by weight changes and positioning errors. This effect was visible in the difference dose plans and the global 2D gamma map with the gamma criterion of 3%/3mm and a threshold of 25%. An example can be seen in Fig. 14 for a patient with a GPR in the upper middle range (P1.2), a patient with a GPR in the lower middle range (P1.9), and the one with the poorest GPR (P1.5).

In the gamma maps in Fig. 14 one can see, that most red voxels, that did not meet the gamma criterion (2D, 3%/3mm, threshold: 25%) are located at the periphery of the patient or at the edge of the density transition from the lung to the tumor. The difference dose distribution show that the voxels not meeting the criterion at the periphery of the patient tend to be too low and those at the transition from the lung to the tumor volume tend to be too high. For patient 5 - who had the worst gamma pass rate - additionally the initial dose distribution on the pCT and that on the sCT_1 are displayed. Here, the change of the patient's shape between the acquisition of the pCT and the CBCT is visible. The initial body contour is visualized in white on the sCT and difference image and the daily adjusted one in blue. Also in this patient it is visible that the dose in the sCT tends to be lower at the periphery of the patient and higher at the transition from the lung to the denser tissue of the tumor volume compared to the dose in the pCT. Changes of the patient weight between the pCT and CBCT acquisition might, therefore, decrease the GPR.
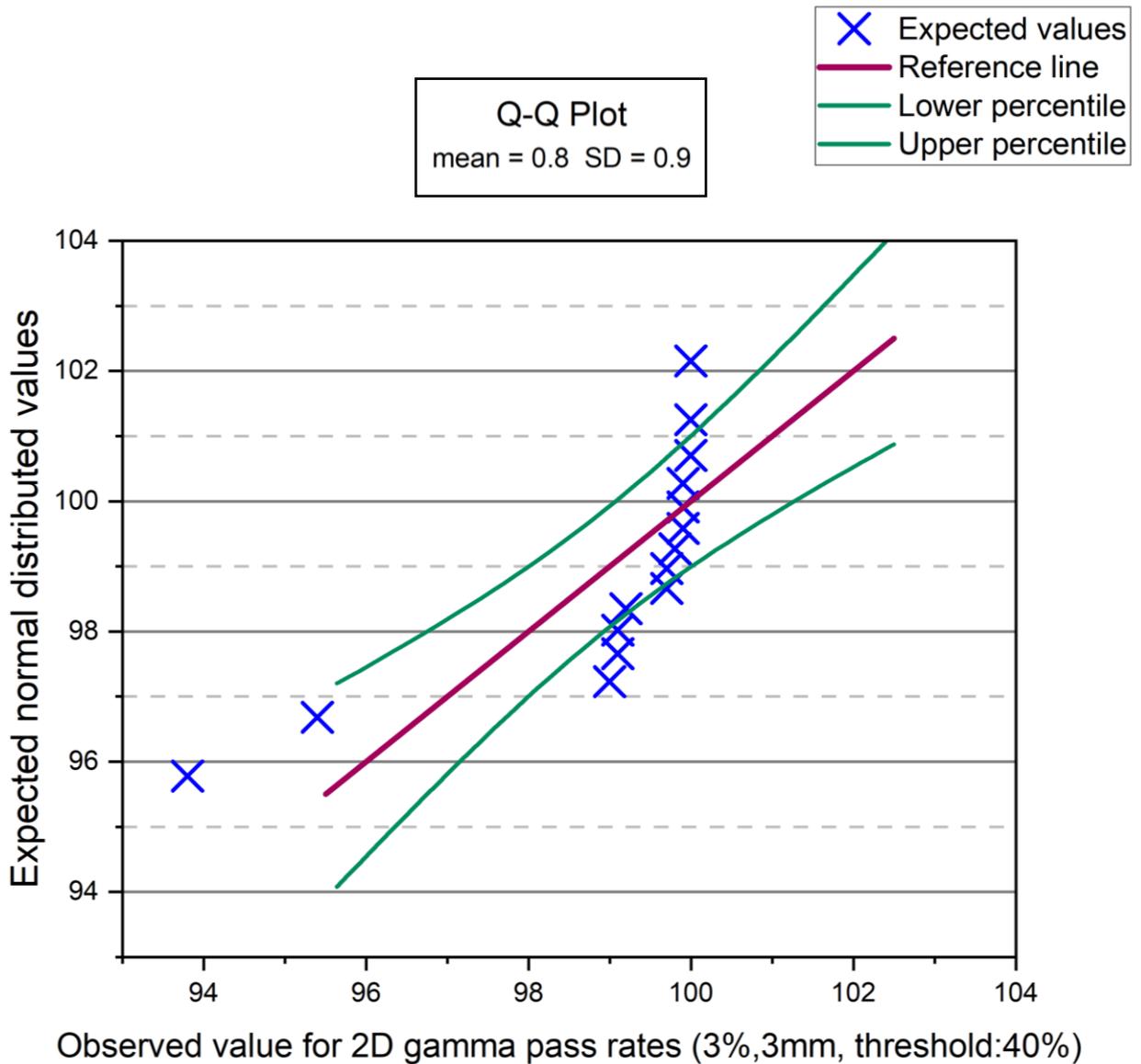
Figure 13 - Quantile-Quantile plots (Q-Q plot) to test if the results of the global 2D Gamma pass rate for the gamma criterion 3 %/3 mm are normally distributed. In the Q-Q plots, the quantiles of the measured values are compared to the theoretical quantiles that the perfectly normally distributed data would have. In the case of a perfect normal distribution, all expected values would be on the reference line. The more they deviate from the reference line, the less the data is normally distributed. Additionally, if all or almost all of the data is within 95% confidence interval (lower and upper percentile), this is an indicator that the data is normally distributed. Here, many data points can be seen outside the confidence interval, indicating non-normally distributed results. With SD=standard deviation.

Table 6 - 2D and 3D gamma pass rates for dose distributions on synthetic CTs compared to the corresponding ones on the treatment planning CT. The results are provided as median±interquartile range (IQR).

	Gamma pass rates (%)							
criterion	2D				3D			
	3%/3mm		2%/2mm		3%/3mm		2%/2mm	
Threshold	25%	40%	25%	40%	10%	25%	10%	25%
Patients								
P1.1	99.0	99.9	96.7	98.9	98.7	99.6	96.8	98.4
P1.2	98.4	99.0	94.6	95.5	84.8	98.4	81.3	94.6
P1.3	99.8	100.0	99.4	99.6	97.7	99.8	96.0	99.4
P1.4	98.8	99.9	97.3	99.0	98.7	98.9	96.6	96.7
P1.5	96.1	93.8	93.7	88.9	94.4	91.6	89.3	84.8
P1.6	98.2	99.8	96.3	99.1	97.0	97.7	94.6	95.5
P1.7	98.6	99.1	96.5	96.9	97.0	98.6	93.4	96.5
P1.8	99.9	100.0	98.6	99.3	96.6	98.3	93.8	96.5
P1.9	97.6	99.1	95.6	95.6	99.6	99.7	98.4	98.5
P1.10	98.0	99.2	95.4	97.3	90.5	91.9	84.0	85.2
P1.11	100.0	100.0	99.5	99.6	97.0	100.0	97.0	99.6
P1.12	99.4	99.7	98.0	98,9	98.9	98.6	96.8	96.2
P1.13	99.7	99.7	98.7	98,6	96.6	99.5	94.3	97.8
P1.14	97.8	95.4	95.1	89.3	96.3	96.2	93.5	92.3
P1.15	99.6	99.9	98.1	98.3	96.9	99.3	95.0	97.5
Median	98.8	99.7	96.7	98.6	97.0	98.6	94.6	96.5
IQR	1.6	0.9	2.8	2.8	1.8	0.7	3.3	3.0
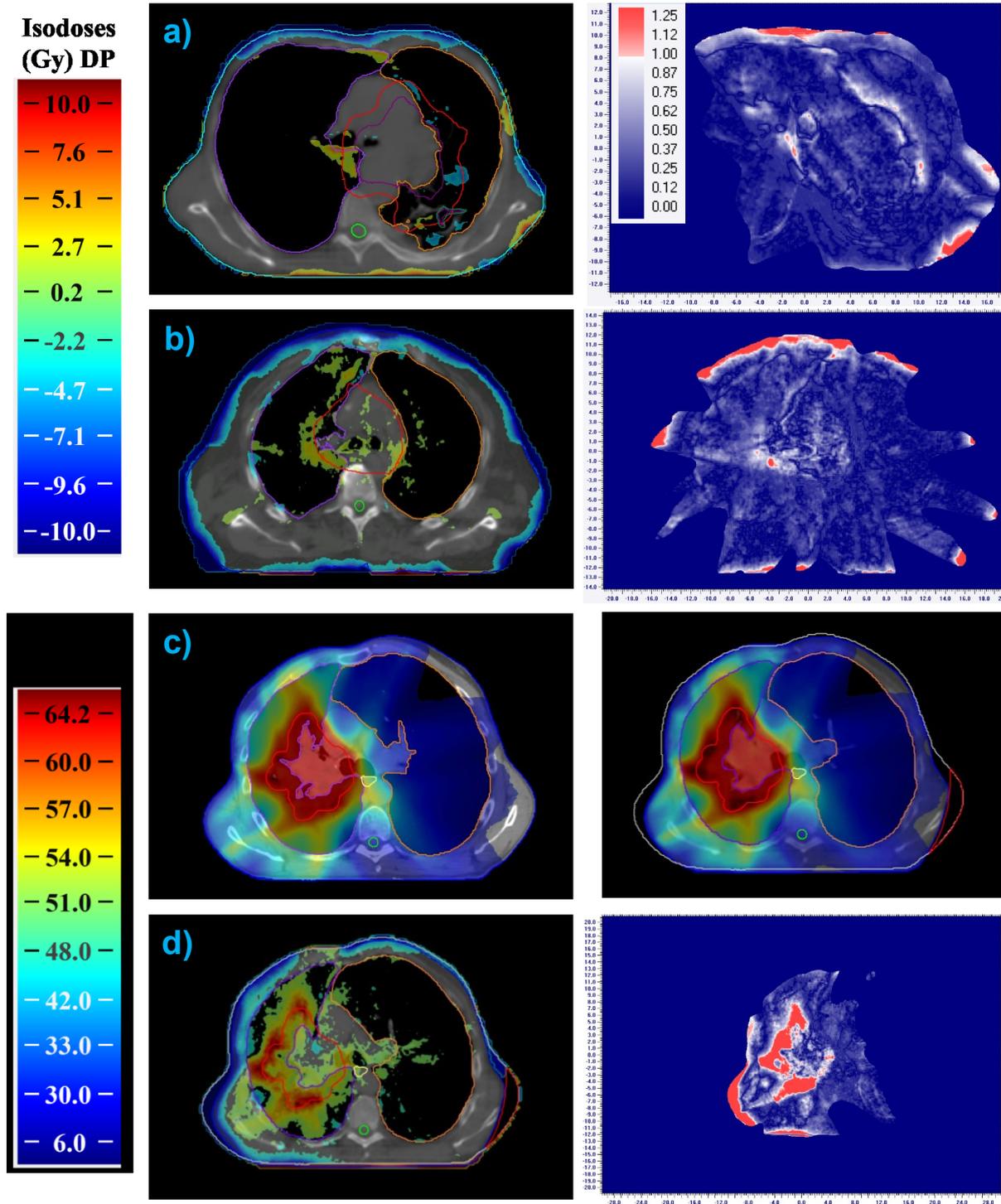
Figure 14 - Difference dose plans (DP) and the corresponding gamma maps (2D, 3 %/3mm, threshold: 25 %) a) for a patient with a gamma pass rate (GPR) in the upper middle range (patient 1.2), b) a patient with a GPR in the lower middle range (patient 1.9) and d) the one with the poorest GPR (patient 1.5). At the edge of the patient, the dose tends to be too cold, and at the transition from the lung to the denser tissue, it tends to be too hot. In c) the initial dose distribution of the treatment plan (TP) on the pCT and that on the sCT$_1$ is shown for patient 1.5 with the worst GPR. The white body contour in the sCT and difference image represents the initial patient outline. The blue one is the adjusted patient outline.

3.2 Image segmentation accuracy analysis

For the image segmentation accuracy analysis all results were normally distributed with p-values of the Kolmogorov-Smirnov test between 0.07 and 1.00. The ipsilateral and contralateral lung were further analyzed as a combined structure, with all p-values of the two-tailed t-test being above 0.05 (range: 0.2-0.8). The results of the DSC, HD, d_{mean}, sensitivity, and specificity of the deformed structures compared to the manually delineated ones are displayed in a boxplot in Fig.15.

Compared to the interobserver DSCs of 0.69-0.88 [60, 107-109] (spinal cord), 0.62-0.82 [60, 107-109] (esophagus), 0.87 [108]-0.94 [60, 107, 109] (heart), and 0.96-0.97 [60, 107, 109] (lungs), the deformed structures are very similar to the manually delineated ones with mean DSCs of 0.82±0.03 (spinal cord), 0.82±0.10(esophagus), 0.84±0.05 (heart), and 0.96±0.02 (lungs). Thus, only the DSC from the heart is lower, which may be due to the poor contrast in the periphery of the CBCT scan where the heart is located. This could also be the reason for the large error bars of the heart's DSC.

The d_{mean} of the spinal cord and esophagus was with 0.9 mm±0.2 mm and 1.1 mm±0.6 mm, in the range of the interobserver variability of 0.7 mm -0.9 mm [60, 109] and 1.1 mm -1.8 mm [60, 109] for the spinal cord and esophagus. The d_{mean} of the heart and lungs exceeded the interobserver d_{mean} values with 6.3 mm±2.3 mm and 2.4 mm±1.3 mm compared to 1.6 mm-3.0 mm [60, 109] and 1.2 mm-1.9 mm [60, 109], respectively. On the one hand, this could be caused by possible anatomical changes between the acquisition of the pCT and the CBCT. On the other hand, the DIR algorithm had problems contouring the last craniocaudal sCT slices of the lung and heart correctly, due to the small FOV in the case of the lung and due to the poor contrast in the periphery of the CBCT scan where the heart is located, resulting in structure inaccuracies. An example for the insufficient contours of the most craniocaudal CT-slices can be seen in Fig. 16. These insufficiently contoured last craniocaudal sCT slices of the lung and heart could also explain the poorer d_{mean}.
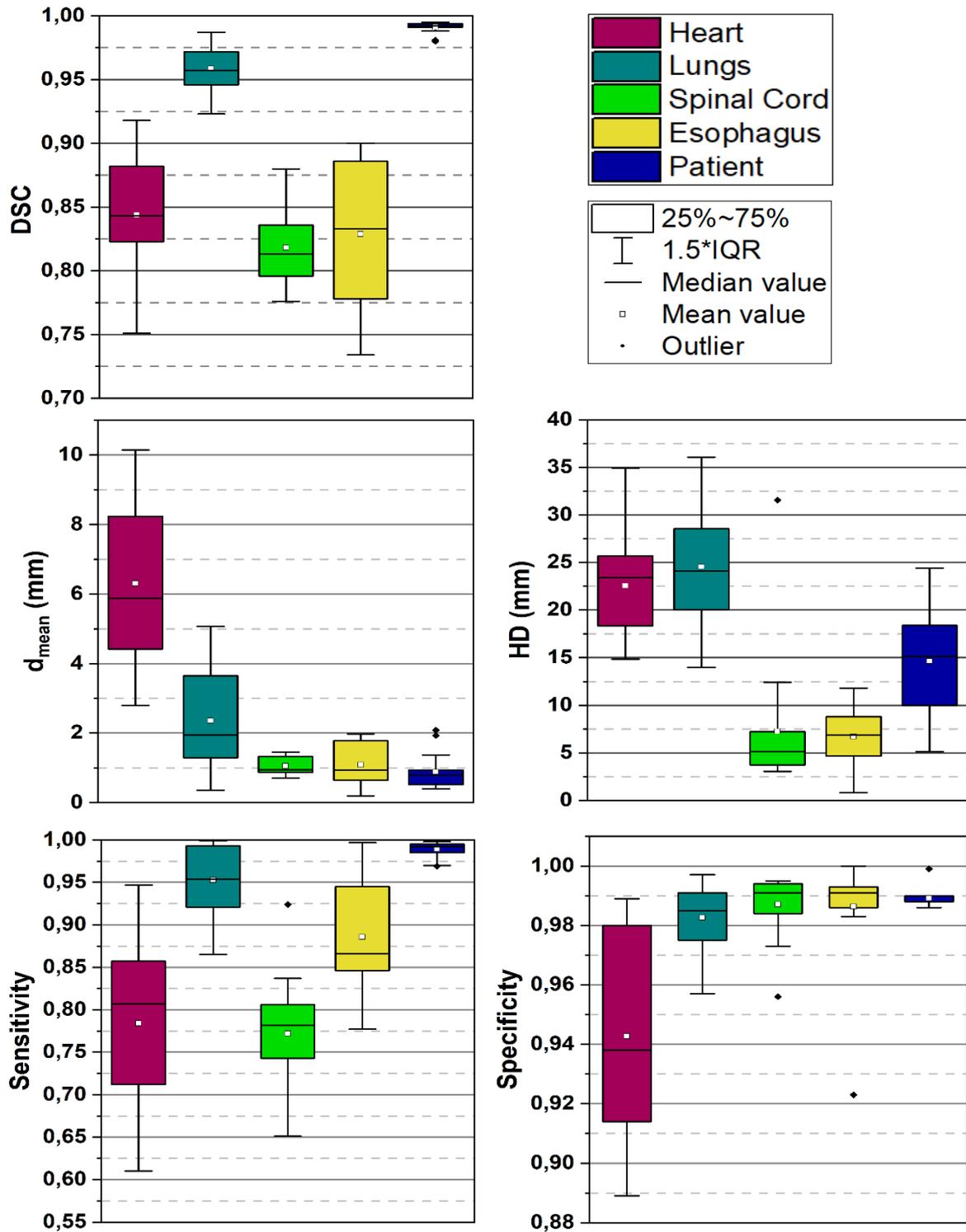
Figure 15 - Dice similarity coefficient (DSC), Hausdorff distance (HD), mean surface distance (d_{mean}), sensitivity and specificity of the heart, lungs, spinal cord, esophagus and the patient outline. The metrics are evaluated between the deformed and manually delineated structures on the synthetic CT. Whiskers represent the data range within 1.5 interquartile ranges (IQR) on the basis of the collective of 15 patients.
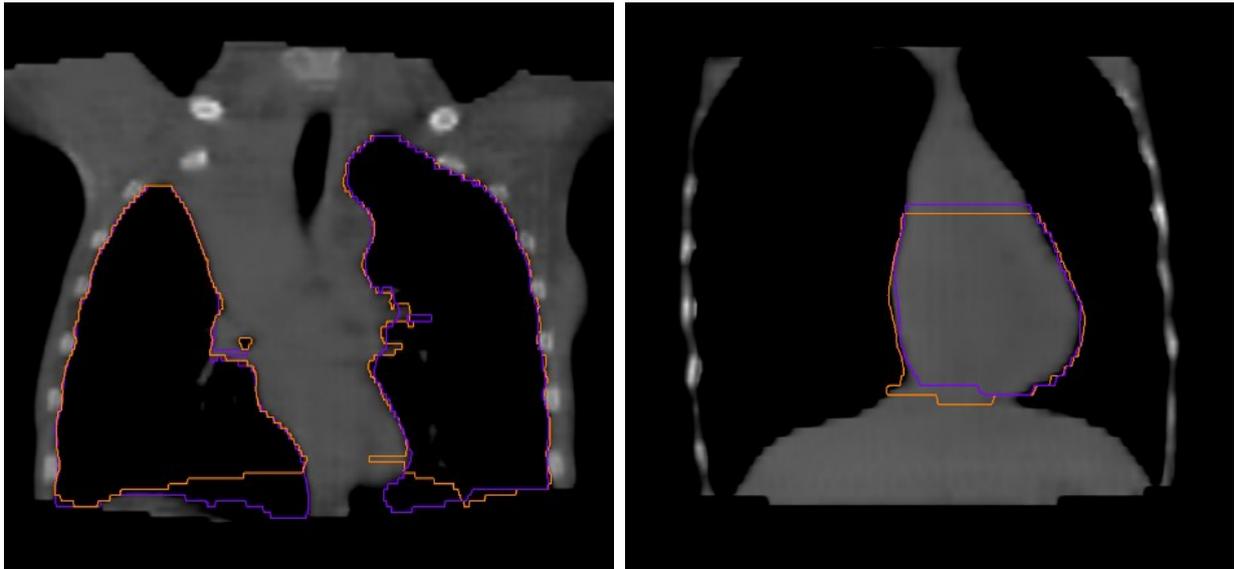
Figure 16 - Insufficient contours of the deformed structures (in orange) for the last craniocaudal slices of the lung (left) and heart (right) compared with manually delineated structures (purple).

The insufficient contours of the last craniocaudally slices influence the HD even more. Therefore the mean HDs of the heart and lungs are also quite high with 22.5 mm±5.5 mm and 24.6 mm±6.2 mm. The mean HDs of the patient outline, spinal cord and the esophagus are lower with 14.6 mm±5.7 mm, 7.2 mm±6.5 mm and 6.7 mm±2.9 mm, respectively.

In all evaluated structures, the specificity is generally higher than the sensitivity, which indicates that the generated structures tend to be larger than the manually delineated structures. The specificity of all structures is higher than 0.9, whereas the sensitivity, on the other hand, shows lower values for the spinal cord and heart. The minimum values for the spinal cord and heart are 0.7 and 0.6, respectively. The mean values for both structures are 0.8±0.1.

The mean time required for the necessary manual corrections of a deformed structure set was 4.9 min±2.2 min. The correction time ranged between 2.1 min and 7.5 min.

3.3 Tumor size variation

The fractional and weekly GTV reduction of each individual and the mean of all patients are all normally distributed. The p-values of the Kolmogorov-Smirnov test were in the range between 0.14 and 1.00. The volume of GTVs decreased for all patients during the treatment period. The mean volume of the GTVs after the whole treatment period was 59.9 %±15.8 %, ranging between 24.0 % and 81.8 %, compared with the GTVs on the sCT_1. After 30 treatment fractions the volume of the GTV were between 24.0 % and 80.6 % with a mean volume of 57.4 %±16.8 %. That equals a mean daily reduction of 0.9 %±0.3 % with a range between 0.4 % and 1.5 % and a mean fractional reduction of 1.4 %±0.6 % ranging between 0.6 % and 2.5 %.

The relative volumes of the GTV over the treatment time are displayed for all patients in Fig. 17 at the top. The GTV progressions of all patients differ from each other and it can be seen that the individual progressions do not follow a clear trend, but rather fluctuate. The strong fluctuations of the volume of the GTV could be caused by image segmentation inaccuracies, image resolution, and actual GTV fluctuations. In the bottom of Fig. 17 the mean volume of the GTVs are shown with the respective SD for each treatment fraction. The strong fluctuation of the daily volume of the GTV within the patient population here becomes visible through the large standard deviation.

The results for the volume changes of the GTV after every 5 treatment fractions as well as the fractional and daily GTV reductions for each patient, the mean, and the SD are shown in Tab. 7. The largest reductions occurred in week 2 with 11.9 %±14.2 % and the least in week 6 with 4.8 %±13.3 %.
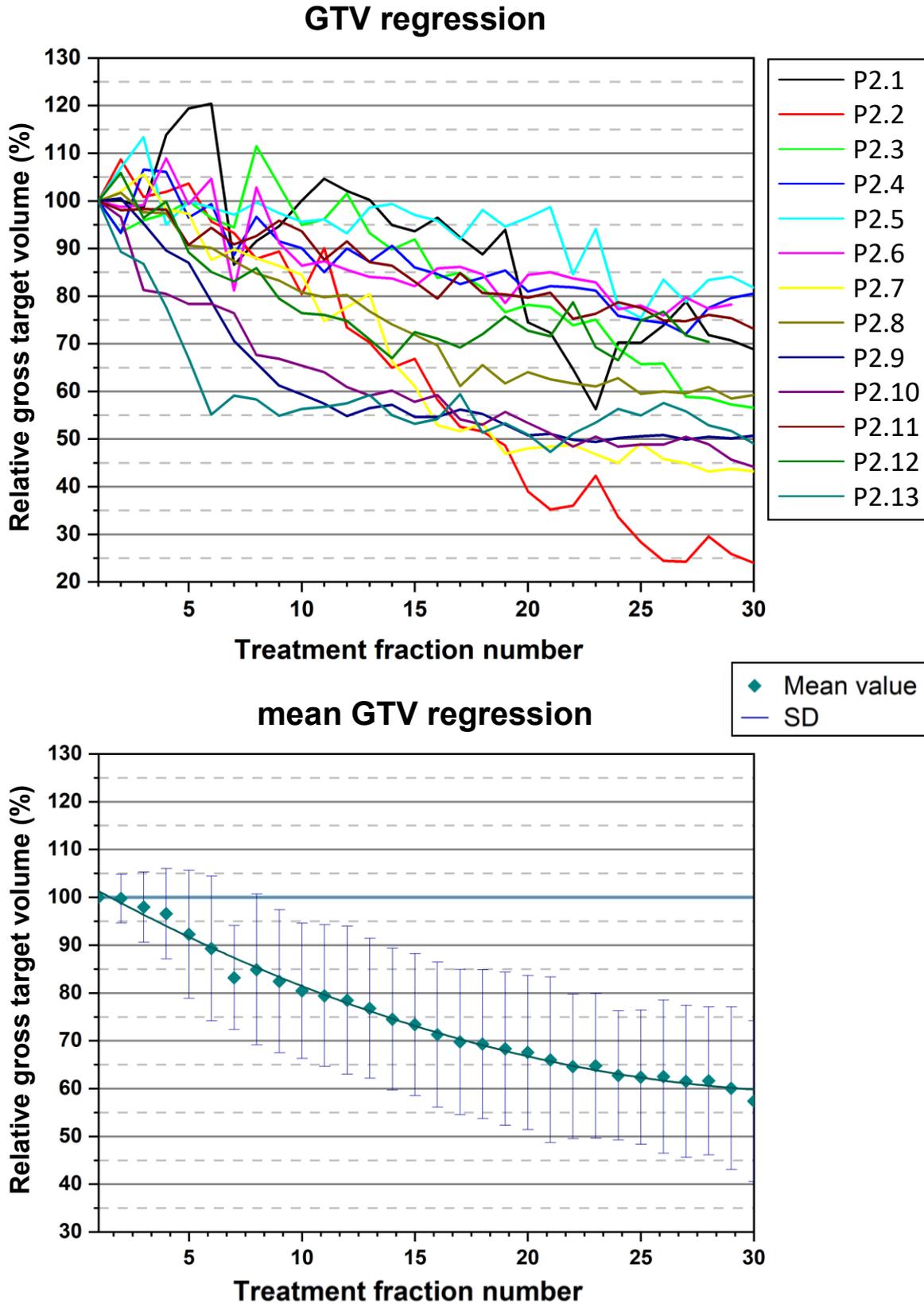
Figure 17 – Relative fraction wise gross tumor volume (GTV) changes. Shown are the GTV dynamics for 13 patients (P2.1-P2.13) with stage III non-small cell lung cancer (at the top) and the averaged GTV dynamics with the standard deviation as error bars (in the bottom).

Table 7 - The relative changes of the gross tumor volume (GTV) after each week of treatment as well as fractional (/fx) and daily (/day) volume changes for the 13 patients, the mean values and the standard deviation (SD).

#			GTV-volume (%)					Reduction (%)	
	5.sCT	**10.sCT**	**15.sCT**	**20.sCT**	**25.sCT**	**30.sCT**	**end**	**/fx**	**/day**
P2.1	119.4	100.1	93.6	74.5	70.2	68.9	68.9	1.0	0.6
P2.2	103.6	80.4	66.8	39.0	28.4	24.0	24.0	2.5	1.5
P2.3	100.0	94.9	92.0	78.2	65.7	56.6	56.6	1.5	0.9
P2.4	96.5	90.0	86.0	81.0	75.0	80.6	80.6	0.7	0.4
P2.5	99.8	95.6	97.0	96.5	75.4	81.8	81.8	0.6	0.4
P2.6	99.3	86.4	82.1	84.5	78.1		78.2	0.8	0.6
P2.7	97.3	84.5	61.1	48.0	49.1	43.2	43.2	1.9	1.1
P2.8	90.6	80.8	71.9	64.1	59.5	59.3	59.3	1.4	1.0
P2.9	87.0	59.4	54.6	50.8	50.6	50.7	50.7	1.6	1.1
P2.10	78.4	65.4	57.8	53.4	48.9	44.2	44.2	1.9	1.2
P2.11	90.7	93.6	83.1	79.7	77.5	73.1	73.1	0.9	0.5
P2.12	89.2	76.4	72.4	72.8	74.9		70.4	1.1	0.7
P2.13	67.0	56.3	53.2	51.0	55.0	49.0	49.0	1.7	1.1
Ø	93.7	81.8	74.7	67.2	62.2	57.4	59.9	1.4	0.9
SD	13.4	14.2	14.9	16.1	14.0	16.8	15.8	0.6	0.3

3.4 Dosimetric benefit of ART for patients with stage III NSCLC

A patient with a mean dose escalation to the PTV of 7.4 Gy (P2.7), which is close to the mean dose escalation, was selected as a representative patient for an exemplary comparison of the dose distributions of the initial treatment plan with those of the IGRT, isotoxic ART, and isoeffective ART scenarios. The V_{20Gy}(lung$_{ipsilateral}$) could be spared in mean by 4.2 % in the isoeffective ART plans. These dose distributions are shown in Fig. 18. In Fig. 18 a) the initial dose distribution on the sCT_1 is displayed and compared with the IGRT, isoeffective ART and isotoxic ART treatment plans on sCT_{30} in Fig. 18 b)-d).

The regressed GTV and slight anatomical changes of the esophagus and lungs can be detected on the sCT_{30} (Fig. 18 b)-d)). The anatomical changes resulted in an isodose of the prescribed 60 Gy in the IGRT treatment plan that is larger than the shrunken PTV (Fig. 18 b)). The isodoses of the IGRT treatment plan seem to be similar but slightly larger than the ones of the initial treatment plan (Fig. 18 a)).

In the isoeffective ART treatment plan (Fig. 18 c)) the isodoses are smaller and more matched to the decreased PTV than in the treatment plan of the IGRT scenario (Fig. 18 b)). The isodoses of the treatment plan of the isoeffective ART scenario (Fig. 18 c)) are also smaller than the ones in the initial treatment plan (Fig. 18 a)).

In the treatment plan of the isotoxic ART scenario (Fig. 18 d)) a clear dose escalation is visible with a maximum dose of 85.1 Gy. The 60 Gy-, 51 Gy- and 42 Gy-isodoses are slightly larger than in the isoeffective ART treatment plan (Fig. 18 c)), but still smaller than in the initial treatment plan (Fig. 18 a)) and the treatment plan of the IGRT scenario (Fig. 18 b)).
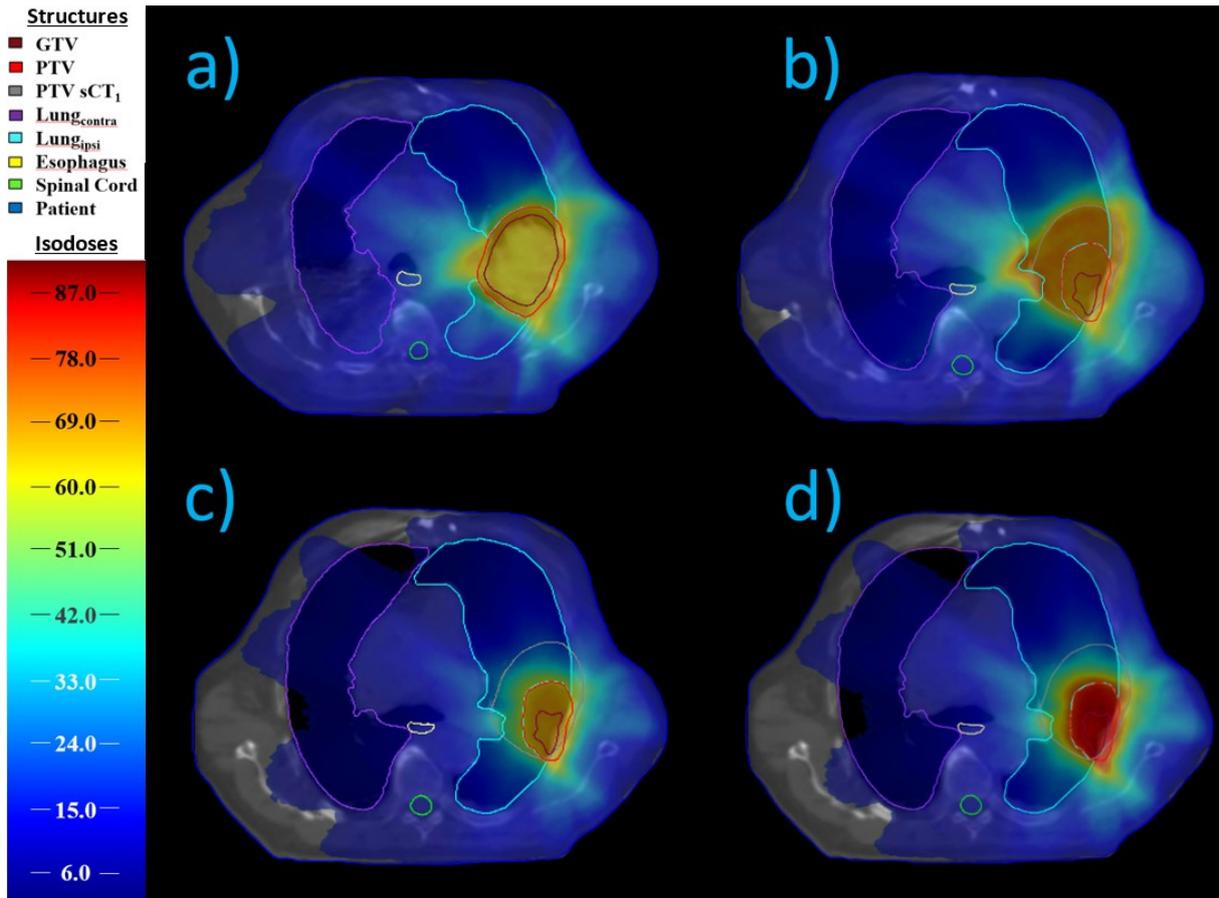
Figure 18 - Dose distributions (a) of the initial treatment plan on the first synthetic CT (sCT) compared to treatment plans on the last sCT b) without adaptive radiotherapy (ART), (c) with isoeffective ART, and (d) with isotoxic ART of a representative patient (P2.7) [81].

None of the DVH parameters of all patients are normally distributed over all treatment fractions of all patients. The p-values of the Kolmogorov-Smirnov test ranged between 0.00 and 0.04. Therefore, the median±IQR are reported. Tab. 8 a) and Tab. 8 b) present the absolute and relative dosimetric values for the following DVH parameters for the three scenarios: $D_{95\%}$(GTV), $D_{95\%}$(PTV), $D_{95\%}$($MD_{initial}$), V_{20Gy}($lung_{ipsilateral}$), $MLD_{ipsilateral}$, V_{5Gy}(heart), MHD, V_{20Gy}($lung_{contralateral}$), $MLD_{contralateral}$, MED, and $D_{0.1\%}$(spinal cord). Furthermore, boxplots of the $D_{95\%}$(GTV), $D_{95\%}$(PTV), and $D_{95\%}$($MD_{initial}$) as well as the V_{20Gy}($lung_{ipsilateral}$), $MLD_{ipsilateral}$, V_{5Gy}(heart), and MHD of the estimated dose accumulation based on average coverage over all treatment fractions of each patient are shown in Fig. 19. The corresponding DVH parameter of the initial treatment plan represents 100 %.

Table 8a - Absolute and relative dosimetric values (median ± interquartile range (IQR)) for the dose volume histogram (DVH) parameters of the target volume for the three scenarios (image guided radiotherapy (IGRT), isoeffective adaptive radiotherapy (ART), and isotoxic ART). With GTV=gross tumor volume, EQD$_2$= the equivalent dose in 2 Gy-fractions, PTV=planning target volume, MD$_{initial}$= microscopic disease of the initial sCT, and TP=treatment plan. Statistically significant differences to the initial treatment plan are marked with an asterisk (*).

Structure	Metric	Median ± IQR (min - max)			
		Initial TP	IGRT	Isoeffective ART	Isotoxic ART
GTV	D$_{95\%}$ (Gy)	60.5 ± 0.9 (59.0-62.3)	60.5 ± 2.0 (41.4-63.3)*	60.4 ± 1.6 (58.0-63.6)*	70.5 ± 9.4 (58.4-122.2)*
	Rel. D$_{95\%}$ (%)	100.0	100.2 ± 1.4 (68.4-103.9)*	99.9 ± 1.9 (96.8-105.0)*	116.0 ± 13.5 (96.3-236.8)*
	EQD$_2$(D$_{95\%}$) (Gy)	60.6 ± 1.1 (58.8-62.8)	60.6 ± 2.4 (38.8-64.0)*	60.5 ± 1.9 (57.6-64.3)*	72.9 ± 11.9 (58.1-147.0)*
	Rel. EQD$_2$(D$_{95\%}$) (%)	100.0	100.3 ± 1.7 (64.1-104.7)*	99.9 ± 2.2 (96.1-106.0)*	119.6 ± 17.0 (95.4-234.0)*
PTV	D$_{95\%}$ (Gy)	60.0 ± 0.0 (56.0-60.0)	57.9 ± 5.8 (16.4-62.1)*	60.0 ± 0.0 (55.8-60.2)	65.5 ± 8.1 (53.5-100.0)*
	Rel. D$_{95\%}$ (%)	100.0	97.1 ± 7.3 (27.4-104.9)*	100.0 ± 0.1 (99.8-100.2)	109.2 ± 12.0 (94.2-221.3)*
	EQD$_2$(D$_{95\%}$) (Gy)	60.0 ± 0.1 (56.0-60.0)	57.7 ± 6.8 (14.1-62.5)*	60.0 ± 0.0 (55.8-60.2)	66.4 ± 9.8 (52.3-113.1)*
	Rel. EQD$_2$(D$_{95\%}$) (%)	100.0	98.8 ± 8.6 (23.5-105.9)*	100.0 ± 0.1 (99.8-100.3)	110.6 ± 14.8 (92.3-188.6)*
MD$_{initial}$	D$_{95\%}$ (Gy)	60.0 ± 0.0 (56.0-60.0)	57.5 ± 7.6 (5.6-61.3)	56.4 ± 9.2 (14.1-60.3)	58.6 ± 8.2 (16.4-73.3)
	Rel. D$_{95\%}$ (%)	100.0	96.5 ± 11.0 (9.3-102.2)	94.3 ± 15.0 (23.5-101.1)	98.8 ± 13.5 (27.3-122.2)
	EQD$_2$(D$_{95\%}$) (Gy)	60.0 ± 0.1 (55.2-60.0)	57.1 ± 8.9 (4.6-61.6)	55.6 ± 10.6 (12.0-60.3)	58.4 ±9.8 (14.0-76.5)
	Rel. EQD$_2$(D$_{95\%}$) (%)	100	93.7 ± 13.0 (7.6-102.6)	91.8 ± 17.3 (20.0-101.3)	98.6 ± 16.1 (23.4-127.5)

Table 8b - Absolute and relative dosimetric values (median±interquartile range (IQR)) for the following DVH parameters: V_{20Gy}(lung$_{ipsilateral}$), ipsilateral mean lung dose (MLD$_{ipsilateral}$), V_{5Gy}(heart), mean heart dose (MHD), V_{20Gy}(lung$_{contralateral}$), contralateral mean lung dose (MLD$_{contralateral}$), mean esophagus dose (MED), and $D_{0.1\%}$(spinal cord). With TP=treatment plan. Statistically significant differences to the initial treatment plan are marked with an asterisk (*).

Structure	Metric	Median ± IQR (min - max)			
		Initial TP	IGRT	Isoeffective ART	Isotoxic ART
Lung$_{ipsilateral}$	MLD (Gy)	14.2 ± 2.4 (7.8-16.9)	14.7 ± 3.0 (7.7-21.8)*	12.3 ± 3.4 (6.6-17.7)*	13.8 ± 2.5 (7.3-20.6)*
	Rel. MLD (%)	100.0	103.6 ± 12.7 (23.5-105.9)*	92.8 ± 10.6 (99.8-100.3)	99.2 ± 3.4 (92.3-188.6)*
	V$_{20Gy}$ (%)	30.2 ± 12.3 (13.0-36.9)	31.3 ± 10.0 (12.7-48.8)*	26.6 ± 14.6 (10.7-40.7)*	30.2 ± 12.4 (13.0-37.0)
	Rel. V$_{20Gy}$ (%)	100.0	105.2±14.7 (74.1-141.8)*	90.7 ± 14.4 (67.9-116.3)*	100.0 ± 0.0 (99.6-100.3)
Lung$_{contralateral}$	MLD (Gy)	6.2 ± 4.4 (3.0-12.9)	6.7 ± 4.8 (2.7-13.7)*	6.0 ± 4.9 (2.2-12.9)*	6.5 ± 4.9 (2.6-13.6)*
	Rel. MLD (%)	100.0	105.1 ± 13.9 (88.1-150.1 *	96.7 ± 10.8 (71.6-124.3)*	102.7 ± 17.6 (73.7-146.8)*
	V$_{20Gy}$ (%)	2.8 ± 8.1 (0-25.4)	3.6 ± 9.0 (0-28.9)	1.7 ± 6.3 (0-29.3)	3.2 ± 10.7 (0-32.3)
	Rel. V$_{20Gy}$ (%)	100	107.4 ± 29.1 (73.9-179.7)	81.2 ± 32.6 (28.0-184.8)	109.2 ± 59.2 (37.6-204.9)
Heart	MHD (Gy)	6.2 ± 5.0 (0.8-9.4)	6.3 ± 5.3 (0.8-15.0)*	5.8 ± 5.4 (0.55-13.0)*	6.2 ± 5.6 (0.8-14.0)
	Rel. MHD (%)	100.0	100.0 ±32.7 (53.6-257.2)*	90.4 ± 20.1 (49.0-164.8)*	100.0 ± 30.1 (46.4-196.1)*
	V$_{5Gy}$ (%)	34.7 ± 24.4 (9.3-59.1)	34.8 ± 26.4 (6.8-100.0)*	30.0 ± 21.7 (4.4-100.0)*	30.8 ± 20.7 (4.5-100.0)*
	Rel. V$_{5Gy}$ (%)	100.0	100.5 ± 23.6 (53.8-187.3)*	90.7 ± 19.3 (48.3-169.3)*	93.0 ± 29.7 (55.3-180.4)*
Esophagus	MED (Gy)	11.9 ± 9.4 (4.0-24.7)	12.5 ± 11.6 (2.9-27.5)*	11.0 ± 10.5 (2.9-25.7)*	11.1 ± 9.4 (2.7-25.1)*
	Rel. MED (%)	100.0	102.7 ±16.1 (51.3-134.6)*	97.2 ± 17.6 (32.0-130.1)*	96.5 ± 22.3 (35.7-147.6)*
Spinal cord	D $_{0.1\%}$ (Gy)	20.3 ± 6.1 (11.3-26.1)	21.3 ± 5.9 (10.6-37.6)*	20.4 ± 6.4 (8.8-29.8)*	21.2 ± 6.8 (7.8-31.0)*
	Rel. D $_{0.1\%}$ (%)	100.0	105.1± 11.1 (84.7-209.8)*	102.1 ± 15.3 (66.1-140.3)*	104.8 ± 16.1 (68.3-188.1)*
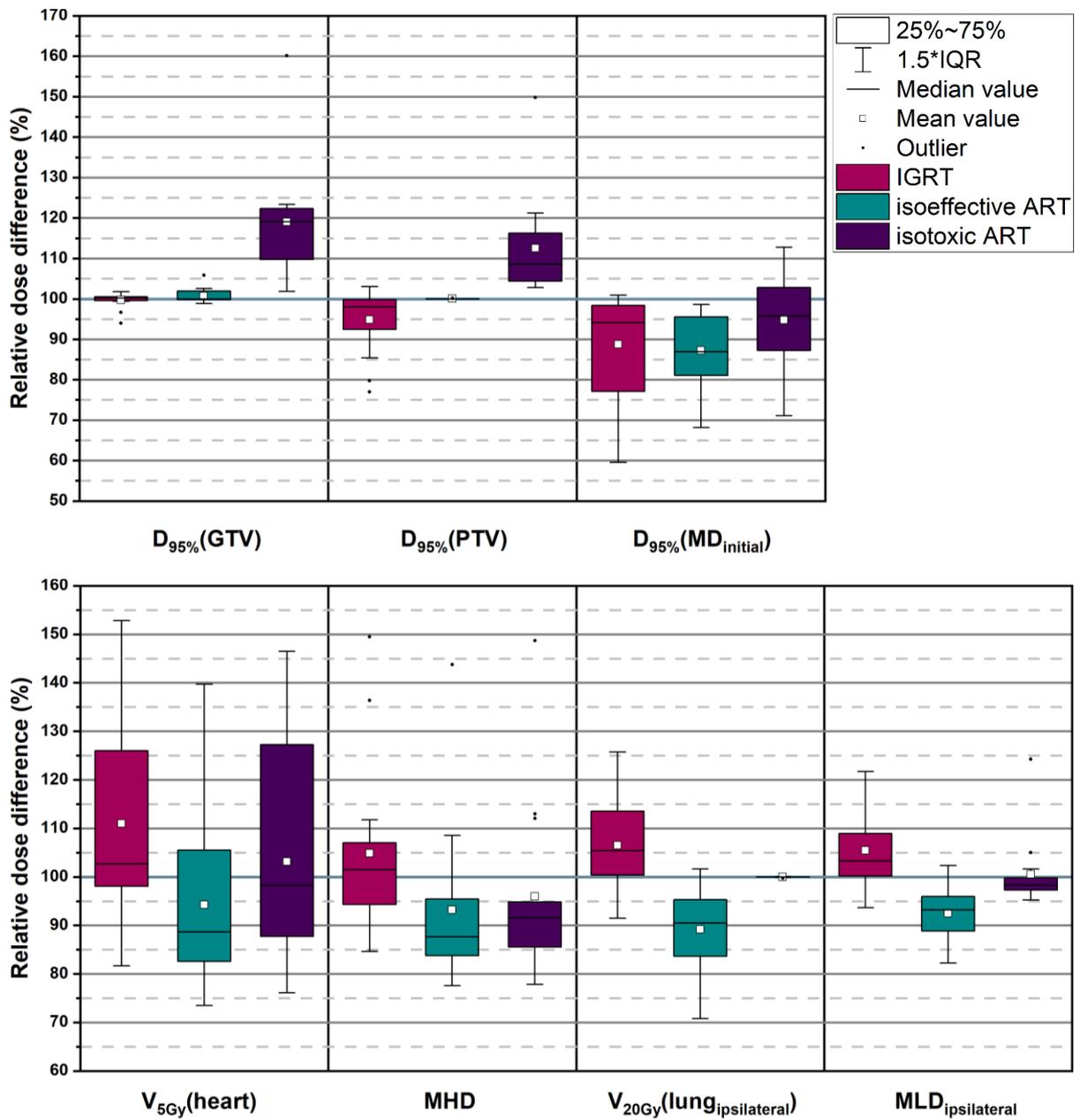
Figure 19 - Dose differences of the estimated dose accumulation based on average values over all treatment fractions of each patient for the following DVH parameters: $D_{95\%}(GTV)$, $D_{95\%}(PTV)$, and $D_{95\%}(MD_{initial})$ as well as the $V_{20Gy}(lung_{ipsilateral})$, $MLD_{ipsilateral}$, $V_{5Gy}(heart)$, and MHD. The corresponding initial dose represents 100 %.

3.4.1 Dosimetric results without ART

While the median GTV coverage remained adequate in the IGRT scenario, the $D_{95\%}$(PTV), $D_{95\%}$(MD$_{initial}$), EQD$_2$($D_{95\%}$(PTV)), and EQD$_2$($D_{95\%}$(MD$_{initial}$)) decreased by 1.6 Gy±4.2 Gy, 2.1 Gy±6.3 Gy, 1.7 Gy±5.0 Gy, and 2.4 Gy±7.2 Gy compared to the initial treatment plans, respectively. The mean $D_{95\%}$(PTV), $D_{95\%}$(MD$_{initial}$), EQD$_2$($D_{95\%}$(PTV)), and EQD$_2$($D_{95\%}$(MD$_{initial}$)) were 3.5 Gy±7.0 Gy, 7.3 Gy±12.1 Gy, 4.0 Gy±7.8 Gy, and 8.0 Gy±12.8 Gy lower than initially planned. The higher decrease of the mean values of these DVH parameters than those of the median values show that there are selected outliers with lower DVH-parameters in the IGRT scenario. The minimum $D_{95\%}$(PTV), $D_{95\%}$(GTV) and $D_{95\%}$(MD$_{initial}$) without ART were 41.4 Gy, 16.4 Gy and 5.6 Gy, respectively. The aforementioned dosimetric constraints (Tab. 4) for the $D_{95\%}$ of the GTV, MD$_{initial}$, and PTV were violated in 21 (5.4 %), 77 (19.9 %), and 108 (27.9 %) treatment fractions, respectively. In the estimated dose accumulation based on the average coverage over all treatment fractions of each patient, the required coverage was not achieved in 1 (7.7 %), 4 (30.8 %), and 5 (38.5 %) patients for the GTV, MD$_{initial}$ and the PTV, respectively.

In the IGRT scenario all analyzed DVH parameters were slightly exceeded compared to the initial treatment plans. The median V_{20Gy}(lung$_{ipsilateral}$), MLD, V_{5Gy}(heart), MHD, MLD$_{contralateral}$, MED, and $D_{0.1\%}$(spinal cord) exceeded the initial dose by 1.1 %±4.4 %, 0.5 Gy±1.8 Gy, 0.2 %±6.7 %, 0.2 Gy±1.2 Gy, 0.4 Gy±0.6 Gy, 0.5 Gy±2.0 Gy, and 0.9 Gy±2.1 Gy, respectively. The mean values showed a larger increase, indicating that a few patients exceeded the parameters more severely. This was particularly noticeable in the mean V_{5Gy}(heart) with a mean increase of 2.5 %±11.1 %. The prior mentioned constraints (Tab. 4) of the V_{20Gy}(lung$_{ipsilateral}$), MLD$_{ipsilateral}$, and the MHD of 37 %, 20 Gy, and 10 Gy were violated in 69 (17.8 %), 6 (1.6 %), and 41 (10.6 %) treatment fractions. This resulted for 3 (23.1 %), 0, and 1 (7.7 %) patients in a violation of the mean DVH values of all treatment fractions per patient for the V_{20Gy}(lung$_{ipsilateral}$), MLD$_{ipsilateral}$, and the MHD, respectively, whereas these dose tolerances were never violated in the initial treatment plans.

3.4.2 Dosimetric results with isoeffective ART

The coverage of the GTV and PTV was restored in the isoeffective ART scenario. Thus, the aforementioned dosimetric constraints (Tab. 4) for the $D_{95\%}$(GTV) and $D_{95\%}$(PTV) were not violated in any treatment fraction. However, the median coverage of the $MD_{initial}$ ($D_{95\%}(MD_{initial})$) and $EQD_2(D_{95\%}(MD_{initial}))$ were 3.7 Gy±8.9 Gy and 5.5 Gy±10.3 Gy lower than initially planned with a minimal dose of 14.1 Gy and 12.0 Gy, respectively, and thus higher than the minimum dose of the IGRT plans. The abovementioned dosimetric constraints (Tab. 4) for the $D_{95\%}$ of the GTV and PTV were also not violated in the estimated dose accumulation based on the mean coverage per patient for any patient. The $D_{95\%}(MD_{initial})$ did not reach the requested coverage in 67 (17.3 %) treatment fractions and in the estimated dose accumulation in 2 (15.4 %) patients.

In the isoeffective ART treatment plans the V_{20Gy}($lung_{ipsilateral}$) and $MLD_{ipsilateral}$ could be spared in median by 3.1 %±3.6 % and 1.4 Gy±1.3 Gy compared to the initial treatment plan. The DVH parameters of the heart were reduced by 2.9 %±6.4 % (V_{5Gy}(heart)) and 0.5 Gy±1.4 Gy (MHD). The MED was decreased by 0.4 Gy±1.6 Gy, whereas the $D_{0.1\%}$(spinal cord) was increased by 0.3 Gy±3.1 Gy. However, the dose increase of the spinal cord was lower than the one in the treatment plans of the IGRT scenario. The dosimetric constraints (Tab. 4) for the V_{20Gy}($lung_{ipsilateral}$), MLD, and the MHD were exceeded for 7 (1.8 %), 0, and 13 (3.4 %) treatment fractions and in the estimated accumulated dose for no patients, respectively.

The dose reduction of the OAR of the isoeffective ART scenario is more prominent comparing the mean values with the ones of the treatment plans of the IGRT approach, with reductions of 4.4 %±5.2 % (V_{20Gy}($lung_{ipsilateral}$)), 1.9 Gy±2.2 Gy ($MLD_{ipsilateral}$), 3.5 %±10.4 % (V_{5Gy}(heart)), 0.8 Gy±2.0 Gy (MHD), 1.2 Gy±1.6 Gy (MED), and 0.5 Gy±3.1 Gy ($D_{0.1\%}$(spinal cord)).

3.4.3 Dosimetric results with isotoxic ART

In the treatment plans of the isotoxic ART scenario a dose escalation (median ±IQR) of 10.0 Gy±8.1 Gy ($D_{95\%}$(GTV)), 6.6 Gy±8.1 Gy ($D_{95\%}$(PTV)), 12.4 Gy±10.3 Gy (EQD_2($D_{95\%}$(GTV))), and 8.1 Gy±13.7 Gy (EQD_2($D_{95\%}$(PTV))), could be achieved. The above mentioned dose constraints (Tab. 4) for $D_{95\%}$(GTV), $D_{95\%}$($MD_{initial}$), and $D_{95\%}$(PTV) were not met in 0, 31 (8.0 %), and 12 (3.1 %) treatment fractions, respectively, in the accumulated dose estimation they were violated for 0, 1 (7.7 %), and 0 patients, respectively. The median $D_{95\%}$($MD_{initial}$) and (EQD_2($D_{95\%}$($MD_{initial}$))) were 1.1 Gy±8.5 Gy and 1.3 Gy±10.1 Gy lower than initially planned, but 5.0 Gy±15.0 Gy and 3.5 Gy±10.1 Gy higher than in the treatment plans of the IGRT scenario, respectively. Compared to the IGRT scenario the median $D_{95\%}$(GTV), $D_{95\%}$(PTV), EQD_2($D_{95\%}$(GTV)) and EQD_2($D_{95\%}$(PTV)) were 10.4 Gy±8.4 Gy, 7.8 Gy±10.4 Gy, 12.8 Gy±10.7 Gy, and 9.4 Gy±12.7 Gy higher, respectively.

There were only outliers of the $D_{95\%}$(GTV) and $D_{95\%}$(PTV) in the direction towards a better coverage. In contrast the $D_{95\%}$($MD_{initial}$) has outliers in both directions, but more towards a worse coverage. The lowest $D_{95\%}$($MD_{initial}$) and EQD_2($D_{95\%}$($MD_{initial}$)) were higher than the one in the IGRT and isoeffective ART scenarios with 16.4 Gy and 14.0 Gy.

In the treatment plans of the isotoxic ART scenario the median contralateral MLD and $D_{0.1\%}$(spinal cord) increased by 0.2 Gy±1.0 Gy and 0.8 Gy±2.7 Gy, with respect to the initial treatment plan. However, they were lower than in the treatment plans of the IGRT scenario by 0.1 Gy±0.9 Gy and 0.1 Gy±3.4 Gy. The median V_{20Gy}(lung$_{ipsilateral}$) remained constant and the median MLD, V_{5Gy}(heart), and MHD were 0.1 Gy±0.4 Gy, 2.2 %±6.0 %, and 0.1 Gy±1.2 Gy lower than initially planned, respectively. Compared to the IGRT trial the median V_{20Gy}(lung$_{ipsilateral}$) was decreased by 1.1 %±4.4 % and the median V_{5Gy}(heart) by 2.4 %±4.5 %. The OAR constraints were violated in 0, 7 (1.8 %), and 27 (7.0 %) treatment fractions for the V_{20Gy}(lung$_{ipsilateral}$), $MLD_{ipsilateral}$, and the MHD, respectively and for the averaged OAR constraints over all treatment fractions of each patient for no patient at all.

3.4.4 DVH parameter progression of a representative patient

For another representative patient (P2.8), whose relative GTV was the closest to the mean GTV at the end of the treatment period with 59.3 %, an analysis of the progression of the dosimetric differences due to the altered anatomy is shown in Fig. 20. The prescription dose for this patient was 60 Gy. The results of all evaluated DVH parameter over the treatment fractions are normally distributed for this patient with p-values of the Kolmogorov-Smirnov test between 0.20 and 1.00. The following DVH parameters are mapped against the individual treatment fractions of the total treatment period for the three scenarios (IGRT, isoeffective ART, and isotoxic ART): $D_{95\%}(GTV)$, $D_{95\%}(PTV)$, $V_{20Gy}(lung_{ipsilateral})$, $MLD_{ipsilateral}$, $V_{5Gy}(heart)$, and MHD. In Fig. 20 the IGRT scenario is shown in blue, the isoeffective ART scenario in green and the isotoxic ART in red. The dashed grey line represents the respective DVH parameters of the initial treatment plan.

One can see the lung sparing ($V_{20Gy}(lung_{ipsilateral})$ and MLD) and target dose escalation ($D_{95\%}(GTV)$ and $D_{95\%}(PTV)$) mentioned above in the isoeffective and isotoxic ART scenarios, respectively. In the IGRT scenario, an up to 1.7 % higher $V_{20Gy}(lung_{ipsilateral})$ and an up to 0.5 Gy higher $MLD_{ipsilateral}$ dose were observed, while the target remained adequately covered across all treatment fractions. In the isoeffective ART scenario the $V_{20Gy}(lung_{ipsilateral})$ and $MLD_{ipsilateral}$ were decreased in mean by 3.0 % and 0.9 Gy with the $D_{95\%}(PTV)$ remaining constant. In contrast, in the isotoxic ART treatment plans, it was possible to increase the mean $D_{95\%}(GTV)$ from 61.3 Gy to 75.6 Gy and the mean $D_{95\%}(PTV)$ to 69.8 Gy without increasing the $V_{20Gy}(lung_{ipsilateral})$. For the $V_{5Gy}(heart)$ and the MHD a strong fluctuation is observable in each scenario.
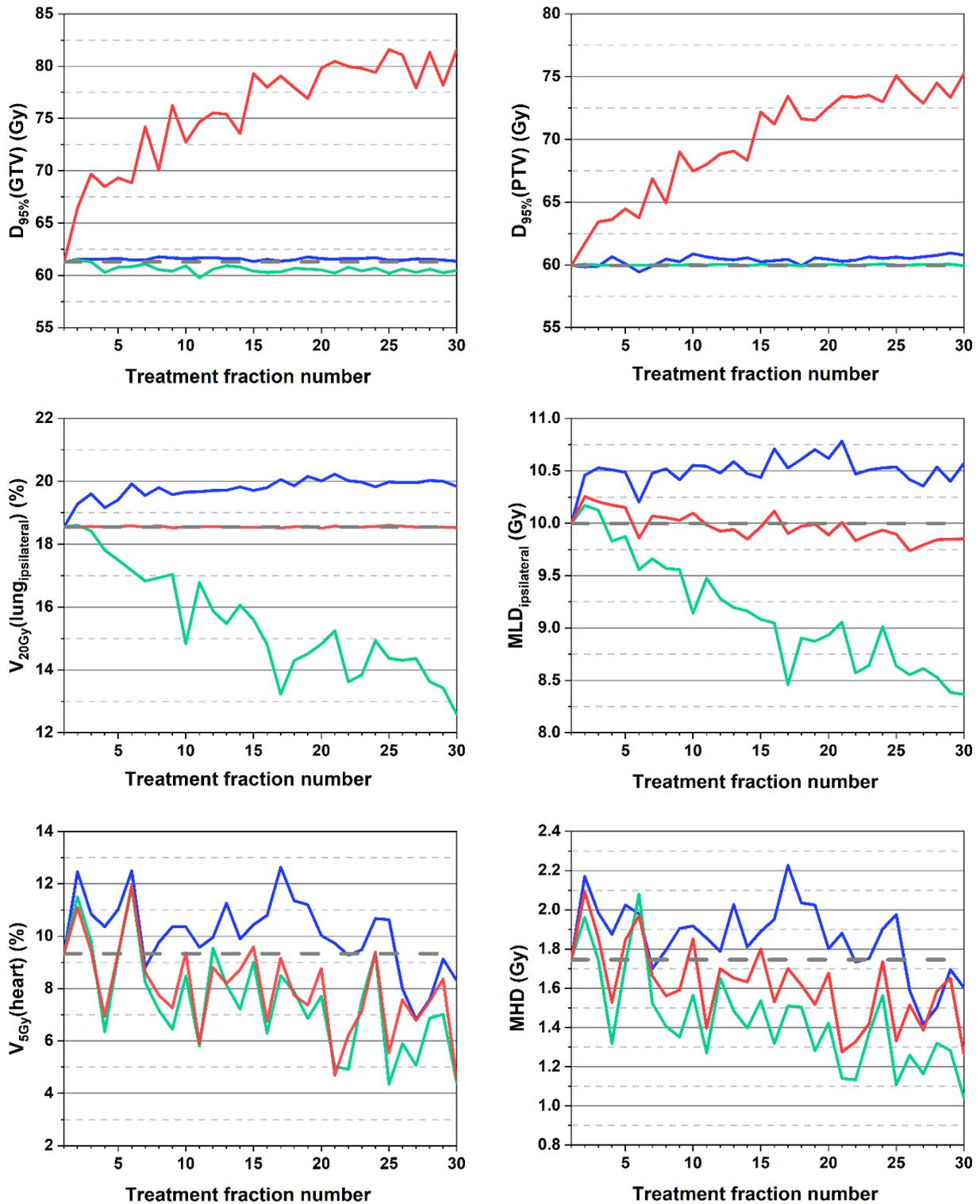
Figure 20 - Analysis of the fraction wise dose volume histogram (DVH) parameters of a representative patient (P2.8). The DVH parameters of the gross tumor volume (GTV) ($D_{95\%}$(GTV)), planning target volume (PTV) ($D_{95\%}$(PTV)), ipsilateral lung (V_{20Gy}(lung$_{ipsilateral}$)) and the heart (V_{5Gy}(heart)) are plotted against treatment fraction number for the three scenarios – with image guided radiotherapy (IGRT) but without adaptive radiotherapy (ART) (blue), isoeffective ART (green) and isotoxic ART (red). The dashed grey line depicts the respective DVH parameters of the initial treatment plan.

3.4.5 DVH parameter correlation with GTV-regression

In the IGRT scenario, the $V_{20\%}$(lung$_{ipsilateral}$) and MLD$_{ipsilateral}$ have an at least moderately strong correlation with the GTV reduction according to the Pearson's correlation coefficient. The Pearson's correlation coefficient of both structures and the reduced GTV was -0.6, which implies an increase of the $V_{20\%}$(lung$_{ipsilateral}$) and MLD$_{ipsilateral}$ with a shrinking GTV. In contrast, the Pearson's correlation coefficient of $V_{20\%}$(lung$_{ipsilateral}$) and MLD$_{ipsilateral}$ for the isoeffective ART scenario were both +0.6, indicating a reduction of the $V_{20\%}$(lung$_{ipsilateral}$) and MLD$_{ipsilateral}$ with a decreasing GTV. In the isotoxic ART scenario, an increased D$_{95\%}$(GTV) was seen with a decreasing GTV, with a Pearson's correlation coefficient of -0.5.

4. Discussion

The developed cycle-GAN algorithm allows for the generation of sCTs based entirely on CBCTs, preserving the patient's actual anatomy. Promising results have been demonstrated in the analysis of image quality, dosimetric and image segmentation accuracy, demonstrating the clinical applicability for ART. Using these sCTs, radiotherapy of patients with stage III NSCLC without the use of adaptive radiotherapy has shown to provide adequate target coverage, but increases the OAR exposure compared to the initial simulated treatment plan. Both isoeffective and isotoxic ART achieved lower OAR doses than the scenario without adaptations, with isotoxic ART also allowing for dose escalation. Ongoing prospective clinical trials will provide clarification on the clinical benefit of this approach.

4.1 Image quality of the generated sCTs
4.1.1 Image uncertainties

Visual comparison of the CBCTs and sCTs with the pCTs indicated that the cycle-GAN algorithm was able to remove many artifacts and almost match the image quality of the reference CT. Moreover, in the IVHs it was evident that the CT numbers of the sCTs matched those of the pCTs better than those of the CBCTs. This can also be observed in the difference images. Image quality also improved significantly with respect to MAE and ME, allowing for a reliable assignment of the CT numbers. In contrast to our study, most studies used a DIR to register the anatomy between the CBCT/sCT with the pCT before comparing the CT numbers. This automatically leads to slightly better MAE and ME values, as it allows to compensate for the geometric changes in the two images. For this reason, the percentage improvement of MAEs is considered as more relevant for the comparison with other studies. The MAE for the total body VOI of several studies was (slightly) lower than our 94.2 HU±31.7 HU with 43.5 HU±6.7 HU [110]; 32.7 HU±7.3 HU [111]; 66.2 HU±8.2 HU [42]; 83 HU±10 HU [53]; 42.5 HU±6.8 HU [112], respectively. However, the mean relative MAE improvement achieved in this study (50.5 %±14.7 %) was within the range of the other studies with 39.8 % [42]-62.1 %[53]. The ME was reported in fewer studies. Nevertheless, it was also lower in these studies than the 29.6 HU±30.0 HU in our study with -2 HU±11 HU [53], 4.3 HU±8.5 HU [113],

and 4.3 HU±7.7 HU [114]. Lenkowitz et al. also obtained better values for the MAE and ME of the lung VOI with 44.5 HU±25.7 HU and -15.9 HU±36.3 HU, respectively [113]. Unexpectedly, Maspero et al. obtained the largest improvement with a 2D cycle-GAN, although they only included 15 patients in the algorithm training [53]. This was the smallest group of training patients compared to the other referenced studies and the algorithm usually improves as the number of training patients increases [115]. Further, they trained three individual algorithms for three different body regions; head and neck, lung, and breast. In addition they trained another algorithm using all training patients for all three regions. This algorithm performed slightly better for the lung than the algorithm trained individually for the lung [53].

4.1.2   Dosimetric accuracy of the generated sCTs

More important, however, is the effect of the slightly different CT number distribution on the dose distribution. In general, the dose calculations on the sCT agreed well with the respective reference dose on the pCT, such that the IQRs of the deviation of all evaluated DVH parameters did not exceed 2.2 % and the mean deviation was not higher than 1.7 %. Except for the $D_{2\%}$(PTV) and V_{30Gy}(lung$_{ipsilateral}$), the mean value was even below 0.9 %. Despite the small deviations, it must be considered that different dose distributions (initial vs. daily) are compared on a possibly slightly different anatomy (initial vs. daily). The changed anatomy also has an influence on the DVH parameters, although the same structures were used, because the density values of the structures change as a result. In addition, the largest deviations came from weight loss of the patient. Thus, the dosimetric deviation may be overestimated. For the $D_{2\%}$(PTV) and V_{4Gy}(heart), there were outliers with deviations of 6.1 % and 3.9 %, respectively. The two outliers both occurred in the same patient who had lost significant weight and whose tumor volume was shifted towards the heart, which may be attributed to the remaining respiratory artifacts despite breath-hold. Lenkowitz et al. trained a GAN algorithm with 40 patients and obtained similar results with deviations for the $D_{2\%}$(PTV), $D_{50\%}$(PTV), $D_{98\%}$(PTV) and $D_{50\%}$(lung) of 4.3 %, -0.1 %, -1.8 %, and 0.7 % on the generated sCT, and 2.9 %, 1.8 %, 0 %, and 0.6 % on a hybrid sCT with an additional bulk density override of the GTV with the electron density relative to water of 1 [113]. Thing et al. used deformable registered and artifact-corrected CBCTs for a dosimetric comparison to the pCT. They had mean dosimetric deviations for the

$D_{2\%}$(PTV), $D_{50\%}$(PTV), $D_{98\%}$(PTV), $D_{0.05cc}$(spinal cord), D_{1cc}(esophagus), MLD, V_{20Gy}(lung), and V_{50Gy}(heart) that were lower than 0.7 % [116], and thus slightly lower compared to our results, where again the deformable image registration might be the reason. With median global GPRs above 98.8 % (2D GPR, 3 %/3 mm, threshold: 25 % and 40 %), 96.7 % (2D GPR, 2 %/2 mm, threshold: 25 % and 40 %), 96.0 % (3D GPR, 3 %/3 mm, threshold: 10 % and 25 %) and 93.4 % (3D GPR, 2 %/2 mm, threshold: 10 % and 25 %) the sCTs provided good results. For all cases except the 2D GPR with the criterion of 2 %/2 mm, threshold: 25 % the mean values were lower than the median values. This demonstrates that there are outliers with a worse gamma passing rate resulting from the fact that some patients have lost weight in between the two image acquisitions. Due to the weight changes and slight positioning differences in most patients the voxels not meeting the gamma criterion are located in the periphery of the patient in the low dose region. Maspero et al. and Lenkowitz et al. reported slightly higher GPR using deformable registration for the 3D gamma analysis with a dose threshold of 10 % with 94.9 %±3.0 % (2 %/2 mm), 98.2 %±1.0 % (3 %/3 mm) [53], 95.5 %±5.9 % (2 %/2 mm) and 98.2 %±4.1 % (3 %/3 mm) [113]. Lenkowitz et al. further improved their GPR using a hybrid sCT with a bulk density overwrite of the GTV with the relative ED of 1 to 96.1 %±5.1 % (2 %/2 mm) and 98.5 %±3.9 % (3%/3 mm) [113].

4.2 Image segmentation accuracy on the generated sCTs

The used DIR algorithm obtained acceptable results in terms of thoracic region contouring metrics. Except for the heart, the mean DSCs were all in the range of the interobserver ranges. The mean DSC of the heart was 0.84±0.05 and the interobserver DSC is in the range of 0.87 [108]-0.94 [60, 107, 109]. The slightly lower DSC of the heart might be caused by the insufficient contrast to the surrounding soft tissue of the heart. The CBCTs contain too many scatter artifacts in the cardiac region which could not be completely removed by the sCT generation with the used cycle-GAN algorithm. A larger number of training patients could potentially improve image quality. In addition, the lower DSC could also arise because of anatomic changes between the sCT and the pCT. Dahiya et al. report of similar mean DSC values for the spinal cord and lungs, a slightly higher DSC for the heart with 0.88±0.08 and a lower DSC for the esophagus with 0.66±0.06 [117]. Furthermore, some of the lung and heart structures generated by the DIR algorithm were insufficient in the last craniocaudally slices (Fig. 16), which

might be caused by the cropped FOV in the cases of the lung structures and the poor contrast in the periphery of the CBCT where the heart is located. This resulted in high d_{mean} and HD values. For the lungs, this effect was more pronounced in cases where the FOV of the CBCTs did not enclose the whole lungs. An option to set a VOI could improve the contouring using a DIR in these cases. The mean HD and d_{mean} values of Dahiya et al. were similar to ours for the spinal cord and esophagus but lower for the lung and heart compared to our results [117]. For all structures in our study, the specificity is larger than the sensitivity, indicating that the generated structures tend to be larger than the manually drawn structures. However, overall, the structures were deemed clinically acceptable after minimal modifications which has markedly accelerated the step of image segmentation.

4.3 General feasibility of ART in NSCLC

Our results can serve as a basis for further improvements of daily imaging protocols and training of sCT models. A larger FOV that encloses the entire lungs would be useful for future dose calculations to accurately evaluate DVH parameters depending on the lung volume. In addition to higher daily imaging doses, this would require an extra training of the cycle-GAN algorithm. Additional training of the algorithm with a larger number of patients could further reduce the remaining shortcomings. Alternatively, several methods could be combined to erase the remaining dosimetric inaccuracies. For example a separate sCT calibration curve could be generated.

With the sCTs reducing artifacts and providing acceptable dosimetric accuracy as well as the accelerated image segmentation provided by the DIR algorithm, an important step has been taken towards implementing ART clinically on a conventional medical linear accelerator capable for IGRT. In addition, the entire ART process is integrated into the research interface of a single planning system, further simplifying its clinical implementation. The required time for the sCT generation never exceeded 30 seconds. Correction of the deformed structures took an average of 4.9 min±2.2 min and never exceeded 7.5 min, thus the time required for sCT generation and structure segmentation never exceeded 8 minutes. This time period is suitable for online ART, but should be further accelerated. Nevertheless, the tested method has demonstrated the feasibility of performing all necessary steps of a possible CBCT-based ART on a conventional linear accelerator within just one software environment.

4.4 Tumor size variation

The GTVs of the patients not only shrunk but also fluctuated, as shown in Fig. 17 at the top. The increase in relative GTV in some treatment fractions could be caused by tumor progression, edema, or inflammation [118]. However, this large variation of the daily volume changes of the GTV cannot be solely explained by this, but is probably also due to contouring inaccuracies caused by the poor contrast and the resolution. The mean GTV regression per day of $0.9\%\pm0.3\%$ was in the lower range of 0.4%-2.6% reported by Sonke et al. [2] and lower than in other research collectives [64, 119, 120]. The GTV recovery rate per day might be influenced by additional breaks due to adverse events in our patient collective. The relative mean size of the GTV after 30 treatment fractions was with $59.9\%\pm15.8\%$ larger than reported by Guckenberger et al. [120] and Zhong et al. [64], but similar to other studies [119, 121, 122]. Dial et al. and Luo et al. only reported a relative GTV at the end of the treatment of 79.9% [74] and 88.3% [123], respectively. The low GTV regression of the two studies could be explained by the different fractionation scheme of Dial et al. [74] and the smaller target volumes in the patient population of Luo et al. [123]. In addition, it was not mentioned whether the patients received a chemotherapy and, if so, whether it was concurrent. Chemotherapy in addition to RT and its timing affects the GTV regression [124]. cCRT has a radio sensitizing effect with the potential for greater tumor regression compared to chemotherapy before or adjuvant to RT or no chemotherapy [122].

4.5 Benefit of ART for patients with stage III NSCLC
4.5.1 Study design, advantages, and limitations

For the evaluation of the benefit of isoeffective and isotoxic ART the initial treatment plans were simulated on the sCT_1 instead of the PTV to further minimize the dosimetric errors. The pCT and all CBCTs were generated in DIBH for all patients. That minimizes inaccuracy due to breathing artifacts. In a study by Brock et al., the MLD and V_{20Gy}(lung) were both 13% lower compared to free breathing. Bainbridge et al. compared DIBH with 4D-CT and yielded significantly lower lung and heart doses with DIBH. The mean relative MHD, MLD, and the V_{21Gy}(lung) were reduced by 11%, 13%, and 10%, respectively in their study [125].

Having one physician performing all the contouring erases the interobserver variations. Furthermore, contouring on the unsorted sCTs limits the influence of the expected results. The heart was difficult to contour and it was not possible to properly distinguish lymph nodes, because of the poor soft tissue contrast. Therefore, the lymph nodes were excluded in this study and the dosimetric outcomes of the heart are more prone to errors and have, because of this, a larger SD. Daily variations in certain DVH parameters, as shown in Fig. 20, reveal the challenge of achieving reproducible treatment scenarios. The variations could be actual volume changes, but they could also be caused by intraobserver variations and resolution problems. In particular, variations in the structure used for dose normalization (ipsilateral lung or PTV) complicate ART and result in dosimetric deviations. This is especially interesting when adaptations are scheduled regularly, but not daily, e.g., once weekly [74].

For dose accumulation in ART, DVFs generated by a DIR are often used to deform the dose distribution of each treatment fraction to the same pCT or sCT. The generation of DVFs is a problem for which there is no unique solution due to too many degrees of freedom [58]. In addition, the overall accuracy of the DIR is highly dependent on image distortions, image resolution, and artifacts [55, 56]. This leads most likely to a dose inaccuracy in dose accumulation using DIR. The inaccuracy is even more pronounced in the case of shrinking structures, as their mass is not preserved [63, 64]. This further complicates dose accumulation in the presence of shrinking tumors [63, 64, 126, 127]. Zhong et al. used different DIR algorithms for ART. For isoeffective ART, they were able to reduce the MLD from 17.3 Gy to 15.2 Gy, 14.5 Gy, or 14.8 Gy depending on the used algorithm. The result is an algorithm based MLD variation of 0.7 Gy (4 %) among the three DIR algorithms for an approximate MLD-reduction of 2 Gy [64]. Therefore, the dose distributions of the individual treatment fractions were not accumulated in this study. The dose distributions of each fraction were evaluated separately. In order to use the standard DVH parameters of the cumulative dose distributions to evaluate the dose distribution of each treatment fraction (e.g., V_{20Gy} of the lung), each fraction was optimized with the total D_{presc}. To evaluate how many DVH constraints were violated, dose accumulation was estimated using the mean of the DVH parameters of all treatment fractions for each patient.

Although daily adaptations of the treatment plans are feasible using modern linear accelerators equipped with CBCT scanners, they are still time-consuming. Furthermore, the dosimetric benefit of ART is decreasing from mid-treatment to weekly to daily treatment plan adaptations [74]. An indicator of the most beneficial treatment fractions for the adaptations or a prognostic tool for OAR toxicities would be advantageous. However, this requires that images with a sufficient image quality are available prior to treatment.

With the regular adjustments of the treatment plans, the GTV-PTV margins can be reduced [2, 33]. If the macroscopic tumor volume is shrinking, it is uncertain whether the MD is shrinking continuously or maintaining its initial volume [2]. In this study, the same safety margins were used for the whole treatment period for all scenarios (IGRT, isoeffective ART, and isotoxic ART). Furthermore, the MD was assumed to stay stationary for decreasing GTVs. Moreover, the initial PTV and not the CTV was used to analyze the dose to the MD to simulate the worst case scenario for ART. Therefore, our results may underestimate the MD dose. In addition, intrafractional anatomical changes were not taken into account in this study. Real-time ART could have even larger dosimetric advantages, but is also linked to an even larger increase of the workload.

Our approach for daily isoeffective and isotoxic ART, was feasible according to an accurate dosimetric analysis. All treatment plans were simulated and evaluated for the same patient population, making it reproducible. Without adaptations, the PTV coverage was slightly reduced and the OAR exposure, on the other hand, increased. The isoeffective scenario spared OAR doses while providing adequate target coverage for all treatment fractions. The isotoxic treatment plans were able to increase the GTV- and PTV-dose without a relevant increase of OAR exposure. The exact dosimetric results without adaptations, with isoeffective ART, and with isotoxic ART are discussed in the next sections.

4.5.2 Dosimetric differences without adaptation

Without any treatment plan adaptations, the GTV coverage remained sufficient, but the median $EQD_2(D_{95\%}(PTV))$ and $EQD_2(D_{95\%}(MD_{initial}))$ were reduced by 1.7 Gy±5.0 Gy and 2.4 Gy±7.2 Gy, respectively. Luo et al. retrospectively recalculated treatment plans on a weekly basis of 24 NSCLC patients using CBCTs. In their patient collective only 81.5 % of the PTVs were actually receiving the D_{presc} compared to 95.6 % in the initial treatment plans [123]. For the OAR all evaluated DVH parameter were exceeded compared to the initial treatment plan. A study by Hoegen et al. showed an increase in lung dose (MLD, V_{20Gy}) and MED for 10 NSCLC patients by daily recalculation using CBCTs without adaptations of the treatment plan, which are comparable to our presented results [119]. In contrast to our study, the $D_{2\%}$(spinal cord) decreased by 1.2 % in the study by Hoegen et al. In our patient population, it increased by 2.7 %±11.2%. The largest difference observed between our results and those of Hoegen et al. is that the MHD of their patient collective increased by 22.7 % and in our patient collective it only increased by 3.0 %±33.3 % [119]. This could be due to the low values of the MHD and the relatively high SD.

The DVH constraints for the tumor coverage restriction of 95 % of the prescribed dose were too low in the estimated dose accumulation by the averaged dose distribution of all treatment fractions per patient in 1 (7.7 %), 4 (30.8 %), and 5 (38.5 %) patients for the $D_{95\%}$(GTV), $D_{95\%}$(MD_{initial}), and $D_{95\%}$(PTV), respectively. The V_{20Gy}(lung$_{ipsilateral}$) and MHD tolerance doses were violated in 3 (23.1 %) and 1 (7.7 %) patients, respectively. In total in 7 out of 13 patients (54 %) violated at least one of the aforementioned dosimetric limits in the estimated accumulated dose. These patients would, therefore, benefit most from ART. Yang et al stated that the need of adaptation is higher for advanced tumor stages (T3-T4). In their study 32 % of patients with NSCLC with a GTV-volume > 147.8 cm³ required an adaptation of the treatment plan [128]. In the patient collective of Agrawal et al. 40 % of the patients with locally advanced NSCLC would benefit from ART [129]. Møller et al. reported adaptation rates of 27 %, but they included NSCLC stage I-II and had less cCRT (78 %). Furthermore, they did not consider an adaptation on a daily basis. In their study, the geometric changes had to persist for at least 3 days, which explains the percentage difference to our results [33]. Appel et al. required a re-plan in 21 % of the patients, but they had a different patient collective which included patients with stage II and patients with small cell lung cancer

(SCLC). Furthermore, they considered a re-plan if the tumor volume changed more than 20 % or a significant target shift occurred [11]. In our patient collective 11 (85 %) of 13 patients experienced a GTV-shrinkage of more than 20 %. However, not only is the total shrinkage important but also the time point of the shrinkage in the treatment period. Woodford et al. stated that ART is required for patients whose GTV shrinkage was > 30 % after 20 fractions [130]. In our case 6 (86 %) of the 7 patients violating at least one dosimetric limit met this thesis. The seventh patients had a shrinkage of 26 % after 20 treatment fractions and in addition tumor movement occurred. When considering for whom ART is most beneficial, one must take into account that the currently known OAR tolerance doses are based solely on initial treatment plans [131]. Thus it is likely that the OAR tolerances for ART will change.

4.5.3   Dosimetric results of isoeffective ART

The isoeffective adaptations of the treatment plans to the altered patient anatomy enabled the PTV coverage restoration while minimizing OAR doses. All patients whose estimated accumulated dose distribution violated a dosimetric tolerance of a DVH parameter maintained within the dosimetric limits with isoeffective ART. The median $V_{20Gy}(\text{lung}_{\text{ipsilateral}})$ and $\text{MLD}_{\text{ipsilateral}}$ were 3.1 %±3.6 % and 1.4 Gy±1.3 Gy lower respectively than initially planned. Hoegen et al. isoeffectively adapted the treatment plan weekly for 10 NSCLC patients. In their results, lung sparing was less evident, which is explainable by the lower adaptation frequency [119]. In their study, the mean $D_{2\%}$ of the spinal cord decreased by 2.5 % and the mean MHD increased by 15.5 %, whereas in our study, the median $D_{0.1\%}$ of the spinal cord increased by 2.1 %± and the median MHD decreased by 10.5 %±23.5 % [119]. The median $D_{0.1\%}$ of the spinal cord in our isoeffective ART scenario exceeded the median of the initial treatment plan, but was still lower than without adaptation. Appel et al. included patients with SCLC and stage II. They obtained a reduction in V_{20Gy} and MLD of the ipsilateral lung of 11.4 % and 11.6 %, respectively, in the cases in which they had to change the simulation technique. When they kept the same simulation technique, the lung sparing was lower by 2.6 % (relative MLD) and 2.7 % (relative $V_{20Gy}(\text{lung}_{\text{ipsilateral}})$). In contrast to our study, in 97 % of the patients, for whom the treatment plan was adapted, only had one adaptation [11]. Dial et al. had lower $\text{MLD}_{\text{ipsilateral}}$ and $V_{20Gy}(\text{lung}_{\text{ipsilateral}})$ sparing, but achieved a higher sparing of the $D_{0.03\%}(\text{spinal cord})$ and MED with daily adaptations of

12 patients with NSCLC [74]. Guckenberger et al. used weekly CTs for treatment plan adaptations in 13 NSCLC patients for simulating ART. Treatment plans were adjusted in week 3 or week 5 or in both weeks. The results of their isoeffective ART with two treatment plan adaptations for the MLD were similar to our results. The reason for the effect of achieving comparable results with only adapting the treatment plan twice was possibly caused by the larger volumes and higher regression rates of the GTV of their patients [120]. In addition, the timing of GTV shrinkage in the treatment period must also be considered. In one of their patients, shrinkage of > 50 % was observed in the first week of the treatment period. In our patient collective, the greatest GTV shrinkage was 34 % after one week [120].

Tumor movement towards the ipsilateral lung and heart resulted in exceeded DVH parameters for the ipsilateral lung and heart in several treatment fractions in two patients. In addition, changes in lung volume caused higher $V_{20Gy}(lung_{ipsilateral})$ in some treatment fractions. To reach the initial PTV-coverage, some dosimetric OAR constraints were violated in these treatment fractions in the isoeffective ART scenario. Increased attention should be paid to this in isoeffective ART and, if the DVH parameters are too high, PTV coverage may have to be reduced.

One uncertain aspect of ART is the possible underdosage of the MD [30]. It is generally a relatively little studied topic. For example, it is not clear whether the MD is shrinking synchronously with the regressing tumor or if it remains stationary [2] and which MD dose is needed to obtain a satisfactory TCP. Some TCP models state that only 70 %-80 % of D_{Presc} is required for MD [103]. In a study by Guckenberger et al. with two treatment plan adaptations, fractionated dose escalation with a constant MLD did not result in an MD underdosage [29]. However, for isoeffective ART, one should carefully examine whether the coverage of the MD is sufficient. In this study the median $EQD_2(D_{95\%}(MD_{initial})$ was 4.9 Gy±10.2 Gy and 1.0 Gy±5.9 Gy lower than in the initial and IGRT treatment plans, respectively. Nevertheless, only 2 (15.4 %) patients violated the required coverage of the $MD_{initial}$ ($D_{95\%} \geq 80 \%$ of the D_{Presc}) in the estimated dose accumulation compared to 4 (30.8 %) patients in the IGRT scenario. Thus, the MD was not underdosed more often in this scenario than without treatment plan adaptations. Nevertheless, special attention should still be paid to the MD dose to avoid an underdosage in general.

In a recent study, Møller et al. evaluated the outcome for patients with NSCLC using isoeffective ART. They adapted the treatment plan when changes visible in the CBCT occurred for three treatment fractions (ART group) and compared these results retrospectively with another group of patients without ART (non-ART). The median MLD and MHD were 4.3 Gy and 3.6 Gy lower, respectively, in the isoeffective ART group than in the non-ART group. It is noteworthy that, unlike in our study, this group used smaller margins for the ART group, which explains their large MLD and MHD reduction [33]. In addition, patients in the non-ART group were less likely to have cCRT. In the ART group, only 20 %, 7 %, and 0.4 % of the patients developed pneumonitis of grade ≥ 2, ≥ 3, or lethal compared with 50 %, 21 %, and 6 % of the non-ART patient population, respectively. In addition, median OS, 2-year OS and PFS after two years improved by 8 month 13 % and 8 %, respectively, in the ART group compared to the control group without ART [33]. Despite the retrospective design of this study, their results clearly show the potential of isoeffective ART. However, new 4D CTs were needed for each adaptation in their study. By using CBCTs already used for patient positioning or CBCT-based sCTs for treatment plan adaptation, the additional dose of 4D CTs could be avoided.

4.5.4   Dosimetric results of isotoxic ART

It is controversial which patients would likely benefit from isotoxic treatment plan adaptations and which upper dose levels are reasonable for a target dose escalation. Some dose escalation studies have found improved outcomes for patients receiving radiotherapy alone or subsequent to chemotherapy with target doses up to 84 Gy [132] and 103 Gy [72]. The possible feasibility of dose escalation has been demonstrated up to 79.2 Gy [75], 83.8 Gy [133], and 94 Gy [17]. In a meta-analysis that included 3795 patients enrolled in 25 trials with a time-corrected EQD_2 range of 36.4 Gy-80.8 Gy Ramroth et al. demonstrated a prolongation of median survival by escalating the target dose for patients without receiving chemotherapy. The opposite was the case with cCRT, which could be caused by higher toxicity levels [15]. In the RTOG-0617 study trial, NSCLC patients treated normofractionated with 60 Gy in 30 treatment fractions were compared to a dose escalation NSCLC patient group receiving 74 Gy in 37 treatment fractions. Both groups underwent a concurrent chemotherapy. The normofractionated group showed better OS rates than the dose escalated patient

group in this study [9]. Increasing the number of treatment fractions rather than the fractional dose results in a prolonged treatment time. Although the possible accelerated repopulation of tumor cells is not sufficient to explain the negative results of the RTOG-0617, it possibly does not provide the most favorable results either [18, 79, 134]. Furthermore, the dose escalated patient group received higher heart-, lung-, and esophagus- doses than the normofractionated patient group [9, 36]. In a meta-analysis with 3600 NSCLC patients receiving cCRT out of 16 trials, Schild et al. found the higher incidence of adverse events of grade III or higher for target doses higher than 60 Gy [135]. A closer examination of the RTOG-0617 patient collective showed a positive effect of dose escalation on the OS of patients with a radioresistant genotype [33]. Therefore, ART may be a strategy for dose escalation in the treatment of selected patients with locally advanced NSCLC, as it can control radiation dose to adjacent OARs. It would be beneficial to investigate first whether a patient is radiosensitive or radioresistant. Sasidharan et al. differentiated between responders and non-responders after 12 treatment fractions with 2 Gy each by the rate of tumor volume shrinkage. Responders continued with the normofractionation of 60 Gy in 30 treatment fractions. Non-responders received a dose escalation of 14 Gy to the PTV on top of the 60 Gy with an additional boost to the remaining active PET-positive volume (average 15 Gy). MED, V_{60Gy}(esophagus), MHD, MLD, V_{20Gy}(lung) were lower, and only the V_{5Gy}(heart) was slightly higher in the non-responder group. No difference was observed between responders and non-responders in the number of toxic events at stage 3 [136]. In this work, unlike in the RTOG-0617 study and the one by Schild et al., a dose escalation with higher fractional doses was chosen for the isotoxic adaptations to avoid possible tumor repopulation due to the prolonged treatment time [18, 79, 134]. Moreover, it was tried not to exceed the OAR doses of the initial treatment plan to avoid the more frequent adverse event occurrence.

In the treatment plans of the isotoxic ART scenario, the median $D_{95\%}$(GTV) increased by 10.0 Gy±8.1 Gy as compared with the one in the initial treatment plans. This GTV dose increase is larger than that of the aforementioned isotoxic ART study by Guckenberger et al. where the mean escalation of $D_{95\%}$(GTV) was 5.7 Gy. Remarkably, they achieved this dose escalation with only two treatment plan adaptations [120]. Compared to the initial treatment plans, the median $EQD_2(D_{95\%}(PTV))$ of the isotoxic plans was higher by 8.1 Gy±13.7 Gy. This is lower compared to the mean dose

escalation of $EQD_2(D_{95\%}(PTV))$ of 13.4 Gy obtained by Weiss et al. as a result of two treatment plan adaptations of 10 NSCLC patients at weeks 2 and 4. The fact that they achieved a larger PTV dose increase with just two adaptations of the treatment plan may be due to our dosimetric limit of $D_{95\%}(PTV)$, the higher GTV regression rate that they reported, the chosen α/β ratio, and in addition, they increased the dose until the initial MLD was exceeded by 1 Gy [137].

During the last years, several studies use a positron emission tomography CT (PET-CT) to define a metabolic tumor volume as a subvolume of the GTV that functions as a boost volume for dose escalation [16, 101, 138]. Møller et al. achieved a mean dose escalation to the metabolic tumor volume and GTV from 66 Gy to 93.9 Gy and 87.5 Gy, respectively without any treatment plan adaptations [16]. In the ongoing RTOG-1106 study trial, patients who underwent a mid-treatment PET-CT used for an individualized target dose escalation up to 80.4 Gy are compared to normofractionated patients. In the escalation patient group with one mid-treatment adaptation a median target dose of 71 Gy was reached. Patients in the dose escalation group achieved a 17.1 % and 10.7 % better primary tumor and in-field 2-year local-regional tumor control, respectively, compared with the normofractionated patient group [101]. Treatment plan adaptations to the remaining PET-positive volumes could yield more favorable results, but are time-consuming and lead to non-negligible additional doses, making PET-CT-based adaptations unsuitable for daily ART [119, 139].

For the isotoxic ART scenario, the $V_{20Gy}(lung_{ipsilateral})$ was kept constant, as this seemed to be the most restrictive dosimetric limit. Reviewing the results retrospectively, this was not always the case. Due to this and the allowed dose normalization ratios from 90 % to 110 % other DVH constraints were violated in some treatment fractions where the $V_{20Gy}(lung_{ipsilateral})$ was not the most limitative OAR constraint. In addition, a strategy for normalization is necessary in cases where the lungs are cropped due to the small FOV of the CBCT. Here, either larger CBCTs should be acquired or an artificial intelligence algorithm that uses prior knowledge to deform the missing region of the lungs from the pCT to the sCT should be used. In addition, the question arises whether it makes sense to normalize to a volume parameter of the ipsilateral lung in the case of daily DVH fluctuations of the ipsilateral lung (Fig. 20) or whether a normalization to a specific volume makes more sense.

Regarding the results of the isotoxic ART scenario, it is important to note that we generated treatment plans with higher fractional doses on CBCT-based sCTs of normofractionated patients. The higher fractional dose might have a different effect on the patient anatomy and the tumor. However, higher fractional doses likely enhance the effect of GTV regression and thus the possibility for dose escalation [64]. Nevertheless, the effects due to the different fractionation are still uncertain.

4.5.5 Potential future implementations

Treatment plan adaptations based on sCTs can be performed on conventional medical linear accelerators that are equipped with CBCT scanners. It is important to note that it may be necessary to train a separate CBCT-to-sCT conversion algorithm, for different CBCT scanners and also for different CBCT presets [80]. In addition, before clinical use of the newly trained algorithm, the image quality of the generated sCTs should be evaluated.

Some new linear accelerators, like the Ethos (Varian Medical Systems, Palo Alto, CA), are directly equipped for online ART, with iterative CBCT reconstructions with a sufficient image quality. Thus no additional CBCT-to-sCT conversion algorithm is needed for adaptations of the treatment plan [140]. The more advanced imaging with the HyperSight (Varian Medical Systems, Palo Alto, CA) enables the acquisition of larger CBCTs with a better resolution and contrast. In addition, the CBCT imaging time has been accelerated, reducing motion artefacts. Furthermore, a new artefact removing algorithm for metal artifacts improves image quality for patients with hip implants [141]. Other options include treatment plan adaptations based on pCTs or PET-CTs, although these involve additional radiation doses [119, 139], or magnetic resonance (MR)-guided adaptations.

MR-guided adaptations do not require ionizing radiation [142, 143]. Furthermore, MR images (MRI) have superior soft tissue contrasts compared to CBCTs and CTs, which simplifies the delineation of OARs and targets [144-147]. However, the improved soft tissue contrast of the MRIs did not lead to a significant improvement of the interobserver variability in lymph node registration [148] and target delineation [149]

compared to CBCTs and CTs, respectively. Another positive aspect of MR-guided ART is that it enables real-time monitoring [145, 150]. Additionally, in combination with functional imaging it can be used to assess the tumor's response to the radiotherapy. This provides the opportunity to adjust the radiotherapy according to biological information [145]. The disadvantages regarding MR-guided ART, on the other hand, are the limited availability of MR-linear accelerators [151] due to the high price [152], the necessary manpower needing to be present throughout the treatment [147], and the duration of the treatment [150, 151, 153]. In addition, the impact of the magnetic field on the dose distribution must be taken into account in MR-guided-ART. The Lorentz force affects secondary electrons at density transitions. It causes the secondary electrons to return to the tissue surface. This can result in increased doses [145, 149].

5. Summary/Outlook

In the current workflow of radiotherapy, possible anatomical changes of the patient, such as tumor motion, tumor volume changes or patient weight changes are neglected. This could lead to differences between the planned and actually delivered dose distribution [5, 29, 30]. In ART, treatment plans are adapted to the patient's actual anatomy [5, 29, 30] based on daily pretreatment images. ART could thereby prevent the discrepancy between the planned and delivered dose, and thereby either reduce the OAR dose or increase the target dose.

Daily calibrated images are needed for treatment plan adaptations, which can be used to accurately calculate the dose distribution. CBCTs show image artifacts due to detector scatter, patient-specific scatter, image lag, and beam hardening. Therefore, there is no definite CT-to-ED calibration, making dose calculations inaccurate. The results presented in this thesis demonstrate the general feasibility of generating sCTs based on daily CBCTs using a cycle-GAN algorithm that preserve the CBCT structures. The generated sCTs have a sufficient image quality, leading to a satisfactory dosimetric accuracy.

These sCTs are also suitable for image segmentation using a DIR algorithm, which accelerates the image segmentation. The DSC of the deformed structures compared to manually delineated ones were within the range of the interobserver variability except for the deformed heart structures, which was in mean 0.03 lower than the minimal interobserver DSC reported in other studies. The generated structures tend to be larger than the manually delineated ones. The HD and d_{mean} of the lung and heart structures were higher than reported in other studies due to insufficient contours in the last craniocaudal slices of the heart and lung. Therefore, all structures generated by the DIR algorithm should be carefully controlled and eventually corrected by an expert physician. The time for generating the sCTs and correcting the structure set never exceeded 8 min. Thus, the generated sCTs and deformed structures demonstrate the theoretical feasibility of CBCT-based ART on conventional linear accelerators within a single software environment.

The reported findings show, that image guidance without adaptations of the treatment plans resulted in decreased but still adequate target coverage, but at the expense of increased OAR exposures. The decreased GTV by 1.4 %±0.6 % per treatment fraction, enabled either a significant sparing of the OARs while maintaining the target coverage or a significant target dose escalation without increasing the $V_{20Gy}(lung_{ipsilateral})$ with daily adaptations of the treatment plan. Both adaptive strategies bring their advantages.

The heart and lung could be spared in median by 3.1%±3.6% $V_{20Gy}(lung_{ipsilateral})$ and 2.9%±6.4% ($V_{5Gy}(heart)$). Lower heart, lung, and esophagus doses are associated with improved OS rates [8, 9, 36, 75, 76]. Especially dose in a subregion at the base of the heart, is significantly associated with OS [76]. Furthermore, toxic events can be avoided by minimizing the OAR doses. The $V_{20Gy}(lung_{ipsilateral})$ and $MLD_{ipsilateral}$ are associated with the dose-limiting toxicity pneumonitis [2]. Cardiac toxicities correlate with higher $V_{5Gy}(heart)$ and MHD [35]. Patients with higher MED also have a higher probability to experience acute esophagitis [12, 13]. In addition, the isoeffective ART scenario increases the probability for patients with locally advanced NSCLC to receive adjuvant immunotherapy [11, 33], which also improves the OS rates [21].

In the isotoxic ART scenario the median $D_{95\%}$ of the GTV and PTV dose could be increased by 10.0Gy±8.1Gy and 6.6Gy±8.1Gy, respectively. Regarding dose escalation, several studies suggested that higher fractional doses increased the TCP and OS for NSCLC patients [69-71]. This was, however, questioned by the results of the RTOG-0617 trial. In the RTOG-0617 trial, normofractionated patients had better outcomes than a high-dose patient group receiving 74 Gy in 2 Gy-fractions. However, in this study, the target dose was increased without treatment plan adaptations. As a result, the exposure of the lungs, heart and esophagus was higher in the high-dose patient group compared to the normofractionated patient group [9, 36]. In particular, the increased dose to the subregion at the base of the heart may explain the poorer results of the high-dose group in the RTOG-0617 study [76]. However, the question concerning for which patients the dose escalation is beneficial up to which dose limit has not yet been fully clarified. Future prospective isotoxic ART studies would be interesting to address this question. In this regard, a distinction must be made between radiotherapy alone or in combination with sequential, and concurrent chemotherapy.

A recommendation for clinical implementation is a recalculation of the current treatment plan in case of detectable changes on the CBCT and the implementation of a threshold of dosimetric changes for the re-planning. For patients whose PTV dose and coverage is sufficient, the primary goal should be OAR-sparing until it is clarified for whom dose escalation to which dose level is beneficial. Only for patients for whom a sufficient PTV coverage could not be achieved in the initial plan, dose escalation should be preferred. Special attention should be paid to the coverage of the MD in the isoeffective approach and to the OARs' DVH parameters which are not used for normalization in the isotoxic approach. For regular treatment plan adaptations, either the presented workflow, a modern adaptive linear accelerator, or an MR linear accelerator should be used. When using the presented workflow, the trained algorithm must first be adapted to the respective CBCT scanner and the dosimetric accuracy must be validated again.

6. References

1. Siegel, R.L., et al., Cancer statistics, 2022. CA Cancer J Clin, 2022. 72(1): p. 7-33.

2. Sonke, J.J. and J. Belderbos, Adaptive radiotherapy for lung cancer. Semin Radiat Oncol, 2010. 20(2): p. 94-106.

3. Howlader, N., et al., The Effect of Advances in Lung-Cancer Treatment on Population Mortality. N Engl J Med, 2020. 383(7): p. 640-649.

4. Liang, J.A., et al., Effectiveness of image-guided radiotherapy for locally advanced lung cancer patients treated with definitive concurrent chemoradiotherapy. Thorac Cancer, 2020. 11(9): p. 2639-2649.

5. Sonke, J.J., M. Aznar, and C. Rasch, Adaptive Radiotherapy for Anatomical Changes. Semin Radiat Oncol, 2019. 29(3): p. 245-257.

6. Kataria, T., et al., Adaptive radiotherapy in lung cancer: dosimetric benefits and clinical outcome. Br J Radiol, 2014. 87(1038): p. 20130643.

7. Yamamoto, T., et al., Changes in Regional Ventilation During Treatment and Dosimetric Advantages of CT Ventilation Image Guided Radiation Therapy for Locally Advanced Lung Cancer. Int J Radiat Oncol Biol Phys, 2018. 102(4): p. 1366-1373.

8. Speirs, C.K., et al., Heart Dose Is an Independent Dosimetric Predictor of Overall Survival in Locally Advanced Non-Small Cell Lung Cancer. J Thorac Oncol, 2017. 12(2): p. 293-301.

9. Bradley, J.D., et al., Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two factorial phase 3 study. The Lancet Oncology, 2015. 16(2): p. 187-199.

10. Wang, K., et al., Heart dosimetric analysis of three types of cardiac toxicity in patients treated on dose-escalation trials for Stage III non-small-cell lung cancer. Radiother Oncol, 2017. 125(2): p. 293-300.

11. Appel, S., et al., Image-guidance triggered adaptive replanning of radiation therapy for locally advanced lung cancer: an evaluation of cases requiring plan adaptation. Br J Radiol, 2020. 93(1105): p. 20190743.

12. Caglar, H.B., M. Othus, and A.M. Allen, Esophagus in-field: a new predictor for esophagitis. Radiother Oncol, 2010. 97(1): p. 48-53.

13. Kaymak Cerkesli, Z.A., E.E. Ozkan, and A. Ozseven, The esophageal dose-volume parameters for predicting Grade I-II acute esophagitis correlated with weight loss and serum albumin decrease in lung cancer radiotherapy. J Cancer Res Ther, 2021. 17(1): p. 94-98.

14. Zhou, C., et al., CT-based radiomics nomogram may predict who can benefit from adaptive radiotherapy in patients with local advanced-NSCLC patients. Radiother Oncol, 2023: p. 109637.

15.	Ramroth, J., et al., Dose and Fractionation in Radiation Therapy of Curative Intent for Non-Small Cell Lung Cancer: Meta-Analysis of Randomized Trials. Int J Radiat Oncol Biol Phys, 2016. 96(4): p. 736-747.

16.	Moller, D.S., et al., Heterogeneous FDG-guided dose-escalation for locally advanced NSCLC (the NARLAL2 trial): Design and early dosimetric results of a randomized, multi-centre phase-III study. Radiother Oncol, 2017. 124(2): p. 311-317.

17.	Belderbos, J.S., et al., Final results of a Phase I/II dose escalation trial in non-small-cell lung cancer using three-dimensional conformal radiotherapy. Int J Radiat Oncol Biol Phys, 2006. 66(1): p. 126-34.

18.	Ma, L., et al., A current review of dose-escalated radiotherapy in locally advanced non-small cell lung cancer. Radiol Oncol, 2019. 53(1): p. 6-14.

19.	Deng, J., et al., Comparison of dosimetric effects of MLC positional errors on VMAT and IMRT plans for SBRT radiotherapy in non-small cell lung cancer. PLoS One, 2022. 17(12): p. e0278422.

20.	Bradley, J.D., et al., Long-Term Results of NRG Oncology RTOG 0617: Standard- Versus High-Dose Chemoradiotherapy With or Without Cetuximab for Unresectable Stage III Non-Small-Cell Lung Cancer. J Clin Oncol, 2020. 38(7): p. 706-714.

21.	Spigel, D.R., et al., Five-Year Survival Outcomes From the PACIFIC Trial: Durvalumab After Chemoradiotherapy in Stage III Non-Small-Cell Lung Cancer. J Clin Oncol, 2022. 40(12): p. 1301-1311.

22.	Spigel, D.R., et al., Five-year survival outcomes with durvalumab after chemoradiotherapy in unresectable stage III NSCLC: An update from the PACIFIC trial. Journal of Clinical Oncology, 2021. 39(15_suppl): p. 8511-8511.

23.	Antonia, S.J., et al., Durvalumab after Chemoradiotherapy in Stage III Non-Small-Cell Lung Cancer. N Engl J Med, 2017. 377(20): p. 1919-1929.

24.	Schmitt, D., et al., Technological quality requirements for stereotactic radiotherapy : Expert review group consensus from the DGMP Working Group for Physics and Technology in Stereotactic Radiotherapy. Strahlenther Onkol, 2020. 196(5): p. 421-443.

25.	Ehrbar, S., 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(1): p. 80-88.

26.	Benedict, S.H., et al., Stereotactic body radiation therapy: the report of AAPM Task Group 101. Med Phys, 2010. 37(8): p. 4078-101.

27.	Boda-Heggemann, J., et al., Deep Inspiration Breath Hold-Based Radiation Therapy: A Clinical Review. Int J Radiat Oncol Biol Phys, 2016. 94(3): p. 478-92.

28.	Casutt, A., et al., Fiducial markers for stereotactic lung radiation therapy: review of the transthoracic, endovascular and endobronchial approaches. Eur Respir Rev, 2022. 31(163).

29.	Lim-Reinders, S., et al., Online Adaptive Radiation Therapy. Int J Radiat Oncol Biol Phys, 2017. 99(4): p. 994-1003.

30. Moller, D.S., et al., Adaptive radiotherapy for advanced lung cancer ensures target coverage and decreases lung dose. Radiother Oncol, 2016. 121(1): p. 32-38.

31. Kaplan, L.P., et al., Cone beam CT based dose calculation in the thorax region. Phys Imaging Radiat Oncol, 2018. 7: p. 45-50.

32. Guckenberger, M., et al., Adaptive radiotherapy for locally advanced non-small-cell lung cancer does not underdose the microscopic disease and has the potential to increase tumor control. Int J Radiat Oncol Biol Phys, 2011. 81(4): p. e275-82.

33. Moller, D.S., et al., Survival benefits for non-small cell lung cancer patients treated with adaptive radiotherapy. Radiother Oncol, 2022.

34. Green, O.L., L.E. Henke, and G.D. Hugo, Practical Clinical Workflows for Online and Offline Adaptive Radiation Therapy. Semin Radiat Oncol, 2019. 29(3): p. 219-227.

35. Wang, K., et al., Cardiac Toxicity After Radiotherapy for Stage III Non-Small-Cell Lung Cancer: Pooled Analysis of Dose-Escalation Trials Delivering 70 to 90 Gy. J Clin Oncol, 2017. 35(13): p. 1387-1394.

36. Vivekanandan, S., et al., The Impact of Cardiac Radiation Dosimetry on Survival After Radiation Therapy for Non-Small Cell Lung Cancer. Int J Radiat Oncol Biol Phys, 2017. 99(1): p. 51-60.

37. Munoz-Unceta, N., et al., Durvalumab in NSCLC: latest evidence and clinical potential. Ther Adv Med Oncol, 2018. 10: p. 1758835918804151.

38. Poludniowski, G.G., P.M. Evans, and S. Webb, Cone beam computed tomography number errors and consequences for radiotherapy planning: an investigation of correction methods. Int J Radiat Oncol Biol Phys, 2012. 84(1): p. e109-14.

39. Disher, B., et al., Correction for 'artificial' electron disequilibrium due to cone-beam CT density errors: implications for on-line adaptive stereotactic body radiation therapy of lung. Phys Med Biol, 2013. 58(12): p. 4157-74.

40. Schulze, R., et al., Artefacts in CBCT: a review. Dentomaxillofac Radiol, 2011. 40(5): p. 265-73.

41. Giacometti, V., A.R. Hounsell, and C.K. McGarry, A review of dose calculation approaches with cone beam CT in photon and proton therapy. Phys Med, 2020. 76: p. 243-276.

42. Qiu, R.L.J., et al., Deep learning-based thoracic CBCT correction with histogram matching. Biomed Phys Eng Express, 2021. 7(6).

43. Fotina, I., et al., Feasibility of CBCT-based dose calculation: comparative analysis of HU adjustment techniques. Radiother Oncol, 2012. 104(2): p. 249-56.

44. de Smet, M., et al., Accuracy of dose calculations on kV cone beam CT images of lung cancer patients. Med Phys, 2016. 43(11): p. 5934.

45. Schröder, L., et al., Evaluating the impact of cone-beam computed tomography scatter mitigation strategies on radiotherapy dose calculation accuracy. Physics and Imaging in Radiation Oncology, 2019. 10: p. 35-40.

References

46.   Hansen, D.C., et al., ScatterNet: A convolutional neural network for cone-beam CT intensity correction. Med Phys, 2018. 45(11): p. 4916-4926.

47.   Jiang, Y., et al., Scatter correction of cone-beam CT using a deep residual convolution neural network (DRCNN). Phys Med Biol, 2019. 64(14): p. 145003.

48.   Li, Y., et al., A preliminary study of using a deep convolution neural network to generate synthesized CT images based on CBCT for adaptive radiotherapy of nasopharyngeal carcinoma. Phys Med Biol, 2019. 64(14): p. 145010.

49.   Harms, J., et al., Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. Med Phys, 2019.

50.   Liang, X., et al., Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy. Phys Med Biol, 2019. 64(12): p. 125002.

51.   Mori, S., R. Hirai, and Y. Sakata, Using a deep neural network for four-dimensional CT artifact reduction in image-guided radiotherapy. Phys Med, 2019. 65: p. 67-75.

52.   Maspero, M., et al., Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. Phys Med Biol, 2018. 63(18): p. 185001.

53.   Maspero, M., et al., A single neural network for cone-beam computed tomography-based radiotherapy of head-and-neck, lung and breast cancer. Physics and Imaging in Radiation Oncology, 2020. 14: p. 24-31.

54.   Sarrut, D., Deformable registration for image-guided radiation therapy. Z Med Phys, 2006. 16(4): p. 285-97.

55.   Schultheiss, T.E., W.A. Tome, and C.G. Orton, Point/counterpoint: it is not appropriate to "deform" dose along with deformable image registration in adaptive radiotherapy. Med Phys, 2012. 39(11): p. 6531-3.

56.   Chetty, I.J. and M. Rosu-Bubulac, Deformable Registration for Dose Accumulation. Semin Radiat Oncol, 2019. 29(3): p. 198-208.

57.   Cardenas, C.E., et al., Advances in Auto-Segmentation. Semin Radiat Oncol, 2019. 29(3): p. 185-197.

58.   Amstutz, F., et al., An approach for estimating dosimetric uncertainties in deformable dose accumulation in pencil beam scanning proton therapy for lung cancer. Phys Med Biol, 2021. 66(10).

59.   Dong, X., et al., Automatic multiorgan segmentation in thorax CT images using U-net-GAN. Med Phys, 2019. 46(5): p. 2157-2168.

60.   Yang, J., et al., Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017. Med Phys, 2018. 45(10): p. 4568-4581.

61.   Buschmann, M., et al., Image guided adaptive external beam radiation therapy for cervix cancer: Evaluation of a clinically implemented plan-of-the-day technique. Z Med Phys, 2018. 28(3): p. 184-195.

62.   Ahunbay, E.E., et al., An on-line replanning scheme for interfractional variations. Med Phys, 2008. 35(8): p. 3607-15.

63. Zhong, H. and I.J. Chetty, Caution Must Be Exercised When Performing Deformable Dose Accumulation for Tumors Undergoing Mass Changes During Fractionated Radiation Therapy. Int J Radiat Oncol Biol Phys, 2017. 97(1): p. 182-183.

64. Zhong, H., et al., Evaluation of adaptive treatment planning for patients with non-small cell lung cancer. Phys Med Biol, 2017. 62(11): p. 4346-4360.

65. Cai, B., et al., A practical implementation of physics quality assurance for photon adaptive radiotherapy. Z Med Phys, 2018. 28(3): p. 211-223.

66. Altman, M.B., et al., A framework for automated contour quality assurance in radiation therapy including adaptive techniques. Phys Med Biol, 2015. 60(13): p. 5199-209.

67. Maraghechi, B., et al., Phantom-based Quality Assurance of a Clinical Dose Accumulation Technique Used in an Online Adaptive Radiation Therapy Platform. Advances in Radiation Oncology, 2023. 8(3): p. 101138.

68. Covington, E.L., et al., Improving treatment plan evaluation with automation. J Appl Clin Med Phys, 2016. 17(6): p. 16-31.

69. Casanova Borca, V., et al., Preliminary evaluation of a novel secondary check tool for intensity modulated radiotherapy treatment planning. Phys Med, 2023. 106: p. 102528.

70. Fontenot, J.D., Evaluation of a novel secondary check tool for intensity-modulated radiotherapy treatment planning. J Appl Clin Med Phys, 2014. 15(5): p. 4990.

71. Glide-Hurst, C.K., et al., Adaptive Radiation Therapy (ART) Strategies and Technical Considerations: A State of the ART Review From NRG Oncology. Int J Radiat Oncol Biol Phys, 2021. 109(4): p. 1054-1075.

72. Kong, F.M., et al., High-dose radiation improved local tumor control and overall survival in patients with inoperable/unresectable non-small-cell lung cancer: long-term results of a radiation dose escalation study. Int J Radiat Oncol Biol Phys, 2005. 63(2): p. 324-33.

73. Machtay, M., et al., Higher biologically effective dose of radiotherapy is associated with improved outcomes for locally advanced non-small cell lung carcinoma treated with chemoradiation: an analysis of the Radiation Therapy Oncology Group. Int J Radiat Oncol Biol Phys, 2012. 82(1): p. 425-34.

74. Dial, C., et al., Benefits of adaptive radiation therapy in lung cancer as a function of replanning frequency. Med Phys, 2016. 43(4): p. 1787.

75. Haslett, K., et al., Isotoxic Intensity Modulated Radiation Therapy in Stage III Non-Small Cell Lung Cancer: A Feasibility Study. Int J Radiat Oncol Biol Phys, 2021. 109(5): p. 1341-1348.

76. McWilliam, A., et al., Demystifying the results of RTOG 0617: Identification of dose sensitive cardiac sub-regions associated with overall survival. J Thorac Oncol, 2023.

77. Kong, F.M., et al., RTOG0617 to Externally Validate Blood Cell ERCC1/2 Genotypic Signature as a Radiosensitivity Biomarker for Both Tumor and Normal Tissue for Individualized Dose Prescription. International Journal of Radiation Oncology*Biology*Physics, 2020. 108(3): p. S2.

78.   Roach, M.C., J.D. Bradley, and C.G. Robinson, Optimizing radiation dose and fractionation for the definitive treatment of locally advanced non-small cell lung cancer. J Thorac Dis, 2018. 10(Suppl 21): p. S2465-S2473.

79.   Zehentmayr, F., et al., Radiation dose escalation with modified fractionation schedules for locally advanced NSCLC: A systematic review. Thorac Cancer, 2020. 11(6): p. 1375-1385.

80.   Eckl, M., et al., Evaluation of a cycle-generative adversarial network-based cone-beam CT to synthetic CT conversion algorithm for adaptive radiation therapy. Phys Med, 2020. 80: p. 308-316.

81.   Hoppen, L., et al., Dosimetric benefits of adaptive radiation therapy for patients with stage III non-small cell lung cancer. Radiat Oncol, 2023. 18(1): p. 34.

82.   Sahiner, B., et al., Deep learning in medical imaging and radiation therapy. Med Phys, 2019. 46(1): p. e1-e36.

83.   Kaji, S. and S. Kida, Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. Radiological Physics and Technology, 2019. 12(3): p. 235-248.

84.   Tian, C., et al., Deep learning on image denoising: An overview. Neural Netw, 2020. 131: p. 251-275.

85.   Zhu, J.-Y., et al., Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. 2017: p. 2242-2251.

86.   Goodfellow, I., et al., Generative adversarial networks. Communications of the ACM, 2020. 63(11): p. 139-144.

87.   Spadea, M.F., et al., Deep learning based synthetic-CT generation in radiotherapy and PET: A review. Med Phys, 2021. 48(11): p. 6537-6566.

88.   Zhu, J., et al. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. in 2017 IEEE International Conference on Computer Vision (ICCV). 2017.

89.   Xu, J., et al., A Deep Convolutional Neural Network for Cone-Beam CT Artefacts Reduction: SU-K-FS4-14. Medical Physics, 2017. 44(6): p. 3011-3012.

90.   Cohen, J. and S. Honari, Distribution Matching Losses Can Hallucinate Features in Medical Image Translation. 2018. p. 529-536.

91.   Wolterink, J.M., et al. Deep MR to CT Synthesis Using Unpaired Data. 2017. Cham: Springer International Publishing.

92.   Kida, S., et al., Visual enhancement of Cone-beam CT by use of CycleGAN. Med Phys, 2020. 47(3): p. 998-1010.

93.   Fan, J., et al., Multimodal image translation via deep learning inference model trained in video domain. BMC Med Imaging, 2022. 22(1): p. 124.

94.   Usui, K., et al., A cycle generative adversarial network for improving the quality of four-dimensional cone-beam computed tomography images. Radiat Oncol, 2022. 17(1): p. 69.

95. van Zijtveld, M., M. Dirkx, and B. Heijmen, Correction of conebeam CT values using a planning CT for derivation of the "dose of the day". Radiother Oncol, 2007. 85(2): p. 195-200.

96. Low, D.A., et al., Dosimetry tools and techniques for IMRT. Med Phys, 2011. 38(3): p. 1313-38.

97. Low, D.A., et al., A technique for the quantitative evaluation of dose distributions. Med Phys, 1998. 25(5): p. 656-61.

98. Elekta AB, Software, C: ABAS: Intra-Patient Deformable Image Registration for Adaptive Radiotherapy – A White Paper. 2014.

99. Ning, M.S., et al., Incidence and Predictors of Pericardial Effusion After Chemoradiation Therapy for Locally Advanced Non-Small Cell Lung Cancer. Int J Radiat Oncol Biol Phys, 2017. 99(1): p. 70-79.

100. Atkins, K.M., et al., Cardiac Radiation Dose, Cardiac Disease, and Mortality in Patients With Lung Cancer. J Am Coll Cardiol, 2019. 73(23): p. 2976-2987.

101. Kong, F.-M.S., et al., NRG-RTOG 1106/ACRIN 6697: A phase IIR trial of standard versus adaptive (mid-treatment PET-based) chemoradiotherapy for stage III NSCLC—Results and comparison to NRG-RTOG 0617 (non-personalized RT dose escalation). Journal of Clinical Oncology, 2021. 39(15_suppl): p. 8548-8548.

102. van Leeuwen, C.M., et al., The alfa and beta of tumours: a review of parameters of the linear-quadratic model, derived from clinical radiotherapy studies. Radiat Oncol, 2018. 13(1): p. 96.

103. van Loon, J., et al., Microscopic disease extension in three dimensions for non-small-cell lung cancer: development of a prediction model using pathology-validated positron emission tomography and computed tomography features. Int J Radiat Oncol Biol Phys, 2012. 82(1): p. 448-56.

104. Tolonen, A., et al., Methodology, clinical applications, and future directions of body composition analysis using computed tomography (CT) images: A review. Eur J Radiol, 2021. 145: p. 109943.

105. Tapio, S., Pathology and biology of radiation-induced cardiac disease. J Radiat Res, 2016. 57(5): p. 439-448.

106. Singh, J.P., et al., Association of Cardiac Resynchronization Therapy With Change in Left Ventricular Ejection Fraction in Patients With Chemotherapy-Induced Cardiomyopathy. JAMA, 2019. 322(18): p. 1799-1805.

107. Tsang, Y., et al., Assessment of contour variability in target volumes and organs at risk in lung cancer radiotherapy. Tech Innov Patient Support Radiat Oncol, 2019. 10: p. 8-12.

108. Sharp, G., et al., Vision 20/20: perspectives on automated image segmentation for radiotherapy. Med Phys, 2014. 41(5): p. 050902.

109. Zhu, J., et al., Preliminary Clinical Study of the Differences Between Interobserver Evaluation and Deep Convolutional Neural Network-Based Segmentation of Multiple Organs at Risk in CT Images of Lung Cancer. Front Oncol, 2019. 9: p. 627.

110. Gao, L., et al., Generating synthetic CT from low-dose cone-beam CT by using generative adversarial networks for adaptive radiotherapy. Radiat Oncol, 2021. 16(1): p. 202.

111. Liu, J., et al., CBCT-based synthetic CT generation using generative adversarial networks with disentangled representation. Quant Imaging Med Surg, 2021. 11(12): p. 4820-4834.

112. Gao, L., et al., Streaking artifact reduction for CBCT-based synthetic CT generation in adaptive radiotherapy. Med Phys, 2022.

113. Lenkowicz, J., et al., A deep learning approach to generate synthetic CT in low field MR-guided radiotherapy for lung cases. Radiother Oncol, 2022. 176: p. 31-38.

114. Thummerer, A., et al., Clinical suitability of deep learning based synthetic CTs for adaptive proton therapy of lung cancer. Med Phys, 2021. 48(12): p. 7673-7684.

115. Rusanov, B., et al., Deep learning methods for enhancing cone-beam CT image quality toward adaptive radiation therapy: A systematic review. Med Phys, 2022. 49(9): p. 6019-6054.

116. Thing, R.S., et al., Accuracy of dose calculation based on artefact corrected Cone Beam CT images of lung cancer patients. Physics and Imaging in Radiation Oncology, 2017. 1: p. 6-11.

117. Dahiya, N., et al., Multitask 3D CBCT-to-CT translation and organs-at-risk segmentation using physics-based data augmentation. Med Phys, 2021. 48(9): p. 5130-5141.

118. Kwint, M., et al., Intra thoracic anatomical changes in lung cancer patients during the course of radiotherapy. Radiother Oncol, 2014. 113(3): p. 392-7.

119. Hoegen, P., et al., Cone-Beam-CT Guided Adaptive Radiotherapy for Locally Advanced Non-small Cell Lung Cancer Enables Quality Assurance and Superior Sparing of Healthy Lung. Front Oncol, 2020. 10: p. 564857.

120. Guckenberger, M., et al., Potential of adaptive radiotherapy to escalate the radiation dose in combined radiochemotherapy for locally advanced non-small cell lung cancer. Int J Radiat Oncol Biol Phys, 2011. 79(3): p. 901-8.

121. Lim, G., et al., Tumor regression and positional changes in non-small cell lung cancer during radical radiotherapy. J Thorac Oncol, 2011. 6(3): p. 531-6.

122. Kang, K.H., et al., A Pilot Study Examining the Prognostic Utility of Tumor Shrinkage on Cone-Beam Computed Tomography (CBCT) for Stage III Locally Advanced Non-Small Cell Lung Cancer Patients Treated with Definitive Chemoradiation. Int J Environ Res Public Health, 2021. 18(6).

123. Luo, J., et al., Study of the Cumulative Dose Between Fractions of Lung Cancer Radiotherapy Based on CT and CBCT Image Deformable Registration Technology. Frontiers in Physics, 2020. 8.

124. van Timmeren, J.E., et al., Tumor regression during radiotherapy for non-small cell lung cancer patients using cone-beam computed tomography images. Strahlenther Onkol, 2020. 196(2): p. 159-171.

125. Bainbridge, H., et al., A Comparison of Isotoxic Dose-escalated Radiotherapy in Lung Cancer with Moderate Deep Inspiration Breath Hold, Mid-ventilation and Internal Target Volume Techniques. Clin Oncol (R Coll Radiol), 2022. 34(3): p. 151-159.

126. Heukelom, J., et al., Differences between planned and delivered dose for head and neck cancer, and their consequences for normal tissue complication probability and treatment adaptation. Radiother Oncol, 2019.

127. Moller, D.S., et al., Adaptive radiotherapy of lung cancer patients with pleural effusion or atelectasis. Radiother Oncol, 2014. 110(3): p. 517-22.

128. Yang, P., et al., Patterns of Local-Regional Failure After Intensity Modulated Radiation Therapy or Passive Scattering Proton Therapy With Concurrent Chemotherapy for Non-Small Cell Lung Cancer. Int J Radiat Oncol Biol Phys, 2019. 103(1): p. 123-131.

129. Agrawal, S., S. Kumar, and A.K. Maurya, Potential for adaptive dose escalation in radiotherapy for patients with locally advanced non-small-cell lung cancer in a low mid income setting. Br J Radiol, 2017. 90(1070): p. 20140234.

130. Woodford, C., et al., Adaptive radiotherapy planning on decreasing gross tumor volumes as seen on megavoltage computed tomography images. Int J Radiat Oncol Biol Phys, 2007. 69(4): p. 1316-22.

131. Marks, L.B., et al., Use of normal tissue complication probability models in the clinic. Int J Radiat Oncol Biol Phys, 2010. 76(3 Suppl): p. S10-9.

132. Rosenzweig, K.E., et al., Results of a phase I dose-escalation study using three-dimensional conformal radiotherapy in the treatment of inoperable nonsmall cell lung carcinoma. Cancer, 2005. 103(10): p. 2118-27.

133. Bradley, J., et al., Toxicity and outcome results of RTOG 9311: a phase I-II dose-escalation study using three-dimensional conformal radiotherapy in patients with inoperable non-small-cell lung carcinoma. Int J Radiat Oncol Biol Phys, 2005. 61(2): p. 318-28.

134. McMillan, M.T., et al., Radiation Treatment Time and Overall Survival in Locally Advanced Non-small Cell Lung Cancer. Int J Radiat Oncol Biol Phys, 2017. 98(5): p. 1142-1152.

135. Schild, S.E., et al., Toxicity Related to Radiotherapy Dose and Targeting Strategy: A Pooled Analysis of Cooperative Group Trials of Combined Modality Therapy for Locally Advanced Non-Small Cell Lung Cancer. J Thorac Oncol, 2019. 14(2): p. 298-303.

136. Sasidharan, B.K., et al., Early PET/CT Response Assessment for Selective FDG PET-Guided Dose Escalation in Locally Advanced NSCLC Patients Enrolled on the FLARE-RT Protocol. International Journal of Radiation Oncology*Biology*Physics, 2018. 102(3): p. e703-e704.

137. Weiss, E., et al., Dose escalation for locally advanced lung cancer using adaptive radiation therapy with simultaneous integrated volume-adapted boost. Int J Radiat Oncol Biol Phys, 2013. 86(3): p. 414-9.

138. Nestle, U., et al., Locally Advanced Non-Small Cell Lung Cancer: Excellent Results of Isotoxic Dose Escalation after FDG-PET-Guided Target Volume Reduction – Results of a Prospective Randomized Multicenter Trial. International Journal of Radiation Oncology*Biology*Physics, 2018. 102(3): p. S16.

139. Smith-Bindman, R., et al., Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Arch Intern Med, 2009. 169(22): p. 2078-86.

140. Mao, W., et al., Evaluation of Auto-Contouring and Dose Distributions for Online Adaptive Radiation Therapy of Patients With Locally Advanced Lung Cancers. Pract Radiat Oncol, 2022. 12(4): p. e329-e338.

141. Systems, V.M., Hypersight a revolution in resolution. 2023.

142. Ahmad, S.B., et al., Evaluation of a commercial MRI Linac based Monte Carlo dose calculation algorithm with GEANT4. Med Phys, 2016. 43(2): p. 894-907.

143. Randall, J.W., et al., Towards Accurate and Precise Image-Guided Radiotherapy: Clinical Applications of the MR-Linac. J Clin Med, 2022. 11(14).

144. Kong, F.M., et al., Effect of Midtreatment PET/CT-Adapted Radiation Therapy With Concurrent Chemotherapy in Patients With Locally Advanced Non-Small-Cell Lung Cancer: A Phase 2 Clinical Trial. JAMA Oncol, 2017. 3(10): p. 1358-1365.

145. Crockett, C.B., et al., Initial Clinical Experience of MR-Guided Radiotherapy for Non-Small Cell Lung Cancer. Front Oncol, 2021. 11: p. 617681.

146. Schiff, J.P., et al., A pilot study of same-day MRI-only simulation and treatment with MR-guided adaptive palliative radiotherapy (MAP-RT). Clinical and Translational Radiation Oncology, 2023. 39: p. 100561.

147. Murray, J. and A.C. Tree, Prostate cancer - Advantages and disadvantages of MR-guided RT. Clin Transl Radiat Oncol, 2019. 18: p. 68-73.

148. Dubec, M., et al., MRI and CBCT for lymph node identification and registration in patients with NSCLC undergoing radical radiotherapy. Radiother Oncol, 2021. 159: p. 112-118.

149. Chin, S., et al., Magnetic resonance-guided radiation therapy: A review. J Med Imaging Radiat Oncol, 2020. 64(1): p. 163-177.

150. Botman, R., et al., The clinical introduction of MR-guided radiation therapy from a RTT perspective. Clin Transl Radiat Oncol, 2019. 18: p. 140-145.

151. Garcia Schuler, H.I., et al., Operating procedures, risk management and challenges during implementation of adaptive and non-adaptive MR-guided radiotherapy: 1-year single-center experience. Radiat Oncol, 2021. 16(1): p. 217.

152. Godley, A., D. Zheng, and Y. Rong, MR-linac is the best modality for lung SBRT. J Appl Clin Med Phys, 2019. 20(6): p. 7-11.

153. Henke, L.E., et al., Stereotactic MR-Guided Online Adaptive Radiation Therapy (SMART) for Ultracentral Thorax Malignancies: Results of a Phase 1 Trial. Adv Radiat Oncol, 2019. 4(1): p. 201-209.

7. Own publications

Journal articles

1) <u>Lea Hoppen</u>, Gustavo R. Sarria, Chung S. Kwok, Judit Boda-Heggemann, Daniel Buergy, Michael Ehmann, Frank A. Giordano and Jens Fleckenstein: Dosimetric benefits of adaptive radiation therapy for patients with stage III non-small cell lung cancer. Radiation Oncology, 18(1): p. 34, 2023

2) Miriam Eckl*, <u>Lea Hoppen</u>*, Gustavo R. Sarria, Judit Boda-Heggemann, Anna Simeonova-Chergou, Volker Steil, Frank A. Giordano, Jens Fleckenstein: Evaluation of a cycle-generative adversarial network-based cone-beam CT to synthetic CT conversion algorithm for adaptive radiation therapy. Physica Medica, 80: p. 308-316, 2020
(*: first and second author contributed equally to this work)

Posters and conference talks:

1) <u>Lea Hoppen</u>, Gustavo R. Sarria, Chung S. Kwok, Judit Boda-Heggemann, Daniel Buergy, Michael Ehmann, Frank A. Giordano, Jens Fleckenstein: Dosimetric benefits of adaptive radiation therapy for patients with non-small cell lung cancer stage III, Poster at the 4th European congress of medical physics (ECMP), Dublin, August 2022

2) Miriam Eckl, <u>Lea Hoppen</u>, Gustavo R. Sarria, Judit Boda-Heggemann, Anna Simeonova-Chergou, Volker Steil, Frank A. Giordano, Jens Fleckenstein: Evaluation of a cycle-generative adversarial network-based synthetic cone-beam CT generation method for adaptive radiation therapy and daily treatment planning. Oral communication at the 3rd ECMP, virtual event, June 2021

3) Miriam Eckl, <u>Lea Hoppen</u>, Gustavo R. Sarria, Volker Steil, Frederik Wenz, Jens Fleckenstein: Clinical applications of adaptive radiotherapy based on CBCT. Talk at the BIR annual radiotherapy and oncology meeting, virtual event, March 2021

4) <u>Lea Hoppen</u>, Miriam Eckl, Gustavo R. Sarria, Judit Boda-Heggemann, Anna Simeonova-Chergou, Volker Steil, Frank A. Giordano, Jens Fleckenstein: Auswertung der Genauigkeit einer künstlichen Intelligenz-basierten Methode zur Erstellung synthetischer CTs für die adaptive Strahlentherapie. Talk at the 51st conference of the Deutsche Gesellschaft für Medizinische Physik (DGMP), virtual event, September 2020

5) Davide Scafa, Jens Fleckenstein, Kerstin Siebenlist, <u>Lea Hoppen</u>, Anna Simeonova-Chergou, Sonja Sudarski, Frederik Wenz, Frank A. Giordano, Judit Boda-Heggemann: Image-guided flattening-filter-free lung SBRT in DIBH: an institutional outcome analysis. Poster at the 25th conference of the Deutschen Gesellschaft für Radioonkologie (DEGRO), Münster, June 2019

6) <u>Lea Hoppen</u>, Anna Simeonova-Chergou, Anika Jahnke, Miriam Eckl, Judit Boda-Heggemann, Volker Steil, Frederik Wenz, Hans-Jörg Wertz, Jens Fleckenstein: Verfahren zur Bestimmung von Elektronendichteverteilungen auf Cone-Beam-CTs (CBCT) für die adaptive Strahlentherapie der Lunge. Poster at the 49th DGMP conference, Nürnberg, September 2018

7) Carsten Ohlmann, Frank Schneider, Anna Arns, Miriam Eckl, <u>Lea Hoppen</u>, Frederik Wenz, Jens Fleckenstein: Experimenteller Vergleich zwischen stereotaktischer Lungenbestrahlung mit volumenmodulierter Strahlentherapie (VMAT) und dynamisch-konformaler Bogenbestrahlung (DCAT) hinsichtlich Bestrahlungszeit und Planqualität. Poster at the 24th conference of the DEGRO, Leipzig, June 2018

8. Curriculum Vitae

Personal information:

Name: Lea Gesa Hoppen

Adress: Waldhofstraße 8

                   68169 Mannheim

Date of birth: 27.09.1990

Place of birth: Nordhorn

Nationality: German

Work experience/professional experience

05/2017 – present   Medical physicist, Department of Radiation Oncology,

                        University Medical Center Mannheim

Education

02/2019 – present   PhD student (doctor scientiarum humanarum)

04/2014 – 03/2016   Heinrich-Heine University Düsseldorf

                        Studies: Medical Physics

                        Completion: Master of Science (01/2016); grade: 1.3

                        Master thesis: „Fluorescence microscopy investigation on single squamous carcinoma cells"; grade: 1.0

10/2010 – 04/2014   Heinrich-Heine University Düsseldorf

                        Studies: Medical Physics

                        Completion: Bachelor of Science (04/2014); grade: 2.1

                        Bachelor thesis: „Binary colloidal mixtures in modulated light fields"; grade: 1.0

08/2003 – 06/2010   Secondary school (Gymnasium Nordhorn)

                        Completion: A levels; grade: 2.1

9. Acknowledgements