Physica Medica 57 (2019) 191–199

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Evaluation of deformable image registration accuracy for CT images of the thorax region

Sebastian Sarudis[a,b,*], Anna Karlsson[a,c], Dan Bibac[d], Jan Nyman[e], Anna Bäck[a,c]

[a] Department of Radiation Physics, Sahlgrenska Academy, University of Gothenburg, 413 90 Göteborg, Sweden
[b] Department of Medical Physics, County Hospital Ryhov, 551 85 Jönköping, Sweden
[c] Department of Therapeutic Radiation Physics, Sahlgrenska University Hospital, 413 46 Göteborg, Sweden
[d] Department of Diagnostic Imaging and Laboratory Medicine, Södra Älvsborgs Hospital, 504 55 Borås, Sweden
[e] Department of Oncology, Sahlgrenska University Hospital, 413 46 Göteborg, Sweden

ARTICLE INFO

Keywords:
Deformable image registration
Breathing motion
Adaptive radiotherapy
Lung cancer

ABSTRACT

Purpose: Evaluate the performance of three commercial deformable image registration (DIR) solutions on computed tomography (CT) image-series of the thorax.
Methods: DIRs were performed on CT image-series of a thorax phantom with tumor inserts and on six 4-dimensional patient CT image-series of the thorax. The center of mass shift (CMS), dice similarity coefficient (DSC) and dose-volume-histogram (DVH) parameters were used to evaluate the accuracy. Dose calculations on deformed patient images were compared to calculations on un-deformed images for the gross tumor volume (GTV) (D_{mean}, $D_{98\%}$), lung (V_{20Gy}, V_{12Gy}), heart and spinal cord ($D_{2\%}$).
Results: Phantom structures with constant volume and shifts ≤30 mm were reproduced with visually acceptable accuracy (DSC ≥ 0.91, CMS ≤ 0.9 mm) for all software solutions. Deformations including volume changes were less accurate with 9/12 DIRs considered visually unacceptable. In patients, organs were reproduced with DSC ≥ 0.83. GTV shifts ≤1.6 cm were reproduced with visually acceptable accuracy by all software while larger shifts resulted in failures for at least one of the software. In total, the best software succeeded in 18/25 DIRs while the worst succeeded in 12/25 DIRs. Visually acceptable DIRs resulted in deviations ≤3.0% of the prescribed dose and ≤3.6% of the total structure volume in the evaluated DVH-parameters.
Conclusions: The take home message from the results of this study is the importance to have a visually acceptable registration. DSC and CMS are not predictive of the associated dose deviation. Visually acceptable DIRs implied dose deviations ≤3.0%.

1. Introduction

External beam radiation therapy may increase survival for patients with lung tumors [1–3]. However, radiation therapy in the thorax region is associated with uncertainties due to respiratory motions that could alter the effect of the treatment from what was planned. This is especially important for modulated treatment techniques such as volumetric modulated arc therapy (VMAT) or intensity modulated proton therapy (IMPT) where the absorbed dose to the target volume is successively built up from different small beam segments covering different parts of the total target volume. The treatment planning and dose calculations for these treatment techniques is usually performed on a 3-dimensional computed tomography (3DCT) image series, which represents the patient anatomy without any information about the internal organ motion. This means that the calculated dose on the planning CT does not take into account organ motions and anatomical changes that may occur during the treatment delivery. In reality, tumors and organs situated in the thorax can move several centimeters during the dose delivery [4–6]. This motion could interplay with the delivery of the modulated treatments, potentially causing effects on the delivered dose to the tumor that are difficult to predict. Furthermore, the patient's anatomy can change throughout the radiation treatment course for example as a response to the treatment [7–9]. Several studies have shown that tumor shrinkage, patient change in weight, and density variations in the lung tissue combined with the physiologic motions of the organs due to respiration may lead to significant differences between the planned and the delivered absorbed dose to the patient if they are not accounted for [10–17].

* Corresponding author.
E-mail addresses: sebastian.sarudis@vgregion.se (S. Sarudis), anna.es.karlsson@vgregion.se (A. Karlsson), dan.bibac@vgregion.se (D. Bibac), jan.nyman@vgregion.se (J. Nyman).

https://doi.org/10.1016/j.ejmp.2018.12.030
Received 16 August 2018; Received in revised form 19 December 2018; Accepted 21 December 2018
Available online 12 January 2019

One potential way to simulate the influence of motion effects caused by respiration is to use a deformable image registration (DIR) software together with images from a 4-dimensional computed tomography (4DCT) scan. A 4DCT scan is composed of several CT series that are collected during different parts of the breathing cycle of a patient. Each CT series corresponds to an individual phase of the breathing cycle. The delivered dose to the patient geometry in each respiration phase can be calculated using the treatment planning system (TPS). By deforming and registering the planning CT image to each of the 4DCT images acquired during the respiration cycle of a patient, the delivered dose within each respiration phase can be summed voxel by voxel back to the planning CT to obtain the total spatial dose distribution to a moving volume.

There are several available software solutions able to perform DIR on medical images, both commercially available, freeware and open source. However, previous studies have reported discrepancies in the deformation accuracy for different algorithms [18–25]. It is therefore important to comprehend the specific limitations for the actual algorithm that is being used in a DIR software before basing any clinical decisions on its results. The SmartAdapt module within the Eclipse TPS, the DIR module in Velocity (both from Varian Medical Systems, Palo Alto, CA) and the RayDeformable module within the RayStation TPS (RaySearch Laboratories, Stockholm, Sweden), are three different commercially available software solutions that can be used to perform DIR between medical image modalities such as 3DCT, 4DCT, CBCT or magnetic resonance imaging (MRI).

This study aims to systematically evaluate and compare the performance of three commercially available DIR software solutions, i.e. SmartAdapt, Velocity and RayDeformable, for deformations performed in the thorax region, and to determine how different parameters such as shift and volume changes of the deformed structures and the slice thickness of the CT image series affect the deformation accuracy. The performance of the software is evaluated both in non-complex geometries in a thorax phantom and in geometries including multi-organ deformations on patient data. In addition to quantitative and visual evaluation of the geometrical deformation accuracy, the impact on dose calculations will also be studied.

2. Material and methods

To evaluate the performance of the deformable image registration (DIR) software solutions, deformations were performed on 3DCT image series of a thorax phantom (CIRS, model 008A) as well as on anatomical 4DCT image series that were acquired during the respiratory cycle of six patients with lung tumors. All the CT image series used in this study were collected on a Toshiba Aquilion Big Bore CT. The accuracy and limitations for each software solution were assessed for five different scenarios:

Phantom study:

1. Deformations of moving tumors without tumor volume changes.
2. Deformations of non-moving tumors with tumor volume changes.
3. Deformations of moving tumors with tumor volume changes.

Patient study:

4. Multi-organ deformations on patient data sets with moving lung tumors.
5. Impact on dose calculations.

2.1. Phantom study

The thorax phantom represents an average human thorax in shape, proportion and composition [26]. The phantom contains a spinal cord and two lungs whereof one of the lungs contains an interchangeable lung equivalent rod connected to a motion actuator box that can be programmed to induce three dimensional motions of the rod. A water equivalent spherical lung tumor of different diameters, ϕ = 10, 20 or 30 mm, can be inserted into the rod. The maximal achievable motion of the tumor insert is 50 mm in the inferior-superior (IS) direction and 11.3 mm in the left-right (LR) and anterior-posterior (AP) directions.

To systemically evaluate how the deformation results were affected by tumor shifts or changes in the tumor volume, 3DCT image series of the thorax phantom were acquired where either the tumor position, the tumor volume or both the position and the volume of the tumor were changed between the image series. For the images containing the tumor with a fixed volume (ϕ = 30 mm), deformations were performed for tumor shifts of 1.0, 5.0, 10.0, 20.0, 30.0, 40.0 and 50.0 mm in the IS direction and 5.7 and 11.3 mm in the LR and AP directions. To obtain changes in the tumor volume without changing the position of the tumor, the lung tumor within the rod was exchanged between the CT image series while the position of the tumor was kept static. The tumor diameter (ϕ) for these cases was varied between ϕ = 10, 20 and 30 mm. The combined effect of both tumor shift and tumor volume change was obtained by either increasing the tumor diameter from ϕ = 10 to 30 mm or decreasing it from ϕ = 30 to 10 mm in combination with a tumor shift of either 0, 10 or 30 mm in the IS direction. To evaluate the impact of the image slice thickness on the deformation results, each CT scan of the thorax phantom was performed twice with slice thickness and reconstruction intervals of 1 and 3 mm respectively.

2.2. Patient study

To study the performance of the algorithms in realistic and more complex geometries, deformations were performed on 4DCT image series of six patients that had received stereotactic body radiation therapy (SBRT) of tumors in the lung. The patients were selected from a cohort of 126 patients that were evaluated to assess typical lung tumor motion and sizes [4]. The observed range of motion for the tumors in these 126 patients was 0–11 mm in the LR and AP directions and 0–53 mm in the IS direction. The patients were divided into three subgroups according to the maximum extension of the tumor diameter. Tumor sizes were defined as follows: small tumors (T1a) $\phi \leq 2.0$ cm, medium tumors (T1b + T2a) $2.0 < \phi \leq 5.0$ cm and large tumors (T2b + T3) $\phi > 5.0$ cm. Two patients from each subgroup were selected for this study.

For each patient, the gross tumor volume (GTV) and the organs at risk (heart, ipsilateral lung and spinal cord) were delineated in the 4DCT images for different respiration phases using the delineation tools in the Eclipse treatment planning system v. 11 (Varian Medical Systems). Furthermore, a planning target volume (PTV) was delineated on the planning CT for each patient. The PTV was derived using the 4DCT-information of the tumor motion to delineate the 4DGTV, i.e. a structure that includes the GTV in all the respiration phases, on which a margin of 3 mm was added to obtain the Internal Target Volume (ITV) and an additional margin of 5 mm was used as the ITV – PTV margin. The diameter of the GTV and volume of the organs at risk for the included patients are presented in Table 1 together with the maximum center of mass shift (CMS) of the GTV between the different respiration phases of the 4DCT. The 4DCT image series were acquired with an image slice thickness of 1 mm and the reconstructed image slice thickness was 2 mm.

2.3. Deformable image registration

For each case, the CT image series were combined in pairs, a source and a target image series. The image series containing the most inferior position of the GTV was selected as the source while the image series containing the most superior position of the GTV was selected as the target. For each combination, the source image series was deformed to the target image series. Each deformation was inspected to determine if the result was visually acceptable. The structures within the two images

Table 1
The maximum extension of the GTV diameter, the center of mass shift (CMS) of the GTV and the absolute volume of the organs at risk for the patients included in this study.

Patient #	Max. GTV diameter [mm]	CMS GTV [mm]	Vol heart [cm^3]	Vol ipsilateral lung [cm^3]	Vol spinal cord [cm^3]
1	17	53	860	1201	17
2	20	23	574	647	14
3	25	23	581	1199	20
4	34	7	384	677	11
5	62	16	688	1475	14
6	64	27	658	693	17

series, i.e. source and target, were delineated by the same person in order to minimize the inter-observer variability. In the phantom the structures were delineated by a physicist while in the patients the structures were delineated by a radiologist. For the patient study, the 4DCT phases containing the most extreme positions of the GTV were always selected for the DIR evaluation.

For the deformations in SmartAdapt, the automatic DIR option was used, for Velocity the extended deformable multipass option was used and for RayDeformable the anatomically constrained deformation algorithm (ANACONDA) was used. The algorithms in SmartAdapt and RayDeformable use an optimization problem to minimize the difference in image intensities between the registered images while the algorithm in Velocity uses an elastic B-spline algorithm that uses mutual information as the cost function metric [25,27–29]. Besides the goal of minimizing the image intensity difference in the images, the ANACONDA algorithm in RayDeformable also uses additional penalization terms such as a grid regularization term to ensure that the grid is smooth and invertible and a shape based grid regularization term that ensures that the deformation is anatomically reasonable. No such penalization terms are used in SmartAdapt. For Velocity, a regularization function is used to generate a smooth deformation vector field and to simulate realistic tissue anatomy and physiological movements. The regularization terms used in Velocity and RayDeformable prevents the voxels from moving independently of their neighbors. The software versions used for the DIR were SmartAdapt v11.0, Velocity v.3.2.0 and RayStation v.5.99.0.16.

Prior to the deformable registration, the algorithms required a rigid registration which was obtained using the auto matching function in each software solution. The volume of interest (VOI) for the rigid registrations was set to cover the spinal cord since this part of the phantom and the patients was considered static. The rigid registrations were confirmed visually before they were accepted. For the DIR in SmartAdapt and Velocity, the VOI was set to cover the entire tumor shift with a margin of 2 cm in each direction for the thorax phantom studies and the entire field of view (FOV) of the 4DCT for the patient data sets. RayDeformable does not use the concept of VOI for the deformations, so these were performed for the entire FOV for each case. To evaluate the impact of the selected size of the VOI on the deformation results, the same pair of patient images were also deformed with the VOI size set to cover the moving tumor with a margin of 2 cm with SmartAdapt and Velocity. This was not studied for RayDeformable since it does not use VOIs for deformations.

For deformations performed with Velocity that resulted in a poor visual agreement for the deformed tumor in the phantom or the GTV in the patients, the DIR was repeated with the “stacked VOI-method” according to recommendations from Varian Medical Systems. This means that a first initial DIR was performed with the VOI covering the entire FOV after which a consecutive DIR was performed with a decreased VOI covering only a smaller volume around the GTV to improve the accuracy. In the results section it is specified which of the results were obtained in this way. This type of selection for deformation was not available in SmartAdapt or RayDeformable.

After deformation, the deformed images were visually examined in order to determine if the DIR was considered visually successful or not. Thereafter the deformed image and the deformed structure sets from Velocity and RayStation were exported to the Eclipse TPS. The statistic tool in SmartAdapt was then used to determine the CMS between the deformed structures from all three software solutions and the corresponding structures in the target image series. To obtain a quantitative measure of the conformity between the deformed structures and the corresponding structures in the target image, the dice similarity coefficient (DSC) was calculated for each structure and deformation according to recommendations from the American association of physicists in medicine, task group 132 (AAPM-TG132) [30].

The DSC is defined as [31]:

$$DSC = \frac{2 \bullet (V_{DIR} \cap V_{Target})}{V_{DIR} + V_{Target}} \quad (1)$$

where V_{DIR} is the volume of the deformed structure and V_{Target} is the volume of the corresponding structure in the target image. $V_{DIR} \cap V_{Target}$ was obtained using boolean operators in Eclipse. A DSC value of 1 means a complete overlap between the two volumes while a value of 0 means no overlap in any voxel for the two volumes. To evaluate the reproducibility between different deformations, each DIR was repeated ten times with the same settings and the CMS and DSC were noted.

In order to compare the DIR accuracy with the uncertainty in manual delineations, the studied structures were delineated manually three times by the same person for each patient and in the phantom, and the DSC and CMS between the delineations were noted.

2.4. Dose calculations

To assess the influence of the DIR accuracy on the dose calculations that are performed on deformed CT-images, a comparison was made between the calculated dose on the deformed image series and the calculated dose on the target image series for each patient case. For each comparison, a volumetric modulated arc therapy (VMAT) plan was optimized on the original planning CT of the patient according to local clinical guidelines on target coverage and constraints to the organs at risk. The VMAT-plans were optimized to obtain an inhomogeneous dose distribution with 100% of the prescribed dose covering the PTV and 146.7% to a center point. Additionally, as large volume as possible of the GTV should be given more than 140% of the prescribed dose. The prescribed dose to each patient was different depending on the position of the tumor, with 15 Gy × 3 to the peripherally located tumors (patient 1, 2, 3 and 5), 10 Gy × 4 to tumors with broad chest wall contact (patient 6) and 7 Gy × 8 to centrally located tumors (patient 4). The treatment plans were re-calculated in Eclipse on both the deformed image and the target image for each DIR software and the difference in the calculated dose distributions was assessed for dose-volume-histogram (DVH) parameters for the included structures (mean dose (D_{mean}) and near minimum dose ($D_{98\%}$) for GTV; volume receiving 20 and 12 Gy (V_{20Gy} and V_{12Gy}) for ipsilateral lung; and near maximum dose ($D_{2\%}$) for spinal cord and heart).

3. Results

3.1. Phantom study

In the first phantom study, the ability of the SmartAdapt, Velocity and RayDeformable software solutions to reproduce the shift of a tumor (0–50 mm in the IS direction and 0–11.3 mm in the LR and AP directions) with a constant volume (ϕ = 30 mm) was investigated. Fig. 1 shows the DSC for the comparison between the tumor structure in the deformed image series and the tumor structure in the target image series for ten repetitive deformable registrations of the tumor shifts in the IS direction. The CT image series were acquired with slice

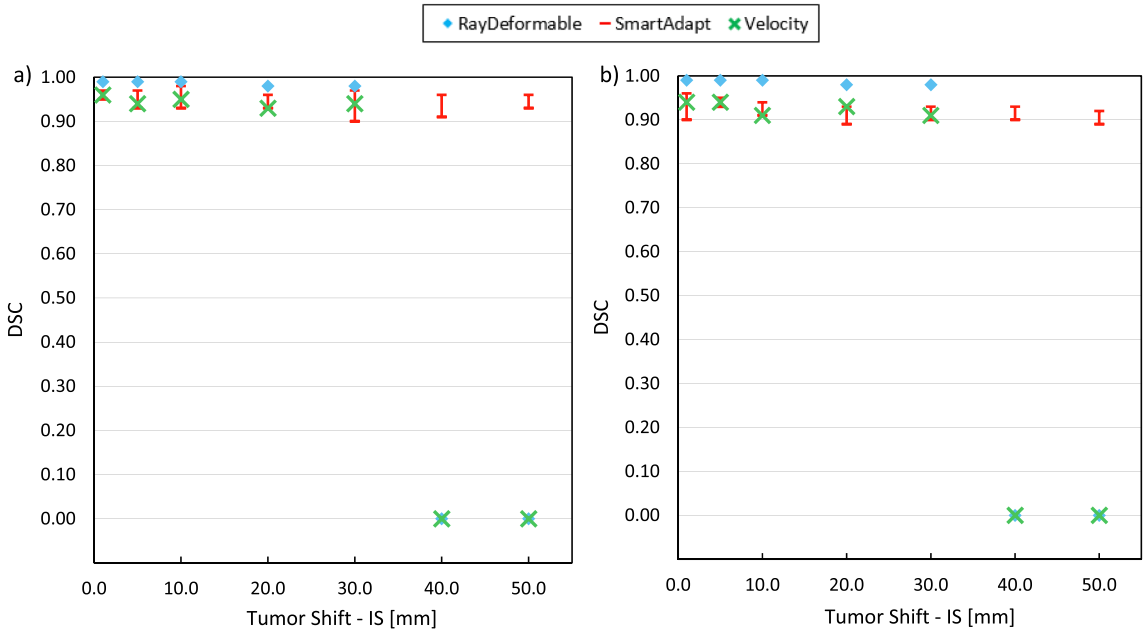

Fig. 1. Dice similarity coefficient (DSC) for the comparison between the tumor structure in the deformed image series and the tumor structure in the target image series for ten repetitive deformations performed on CT image series with slice thickness of 1 mm (a) and 3 mm (b) for a moving tumor in the IS direction with a diameter of 30 mm. The obtained DSC varied between deformations with SmartAdapt (indicated as a red interval illustrating the maximum and minimum DSC) while Velocity and RayDeformable gave the same results for each repetition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

thicknesses of 1 and 3 mm. For DIRs performed with the SmartAdapt software solution, a difference in the results was observed for each deformation that was performed even though all input parameters were the same. This is seen in Figs. 1–4 as the results for SmartAdapt are presented as bars which indicate the maximum and the minimum obtained DSC for the different cases while the results for the Velocity and RayDeformable software solutions are presented as crosses or points since repeated deformations with these algorithms gave the same results each time.

The highest obtained DSC was ≥ 0.92 for all the deformed registrations performed with the SmartAdapt software. For the Velocity and RayDeformable software solutions the DSC was ≥ 0.91 and ≥ 0.98 respectively for shifts up to 30 mm while it was zero for shifts larger than 30 mm. The corresponding CMS was ≤ 0.9 mm in all directions for the SmartAdapt software, while it was ≤ 0.8 mm for the Velocity software and ≤ 0.1 mm for the RayDeformable software for shifts up to 30 mm. The shifts ≥ 30 mm could not be reproduced by Velocity with an automatic DIR (DSC = 0), so the stacked VOI-method was used for these cases. For the 30 mm shift the DSC thereby increased from 0 to ≥ 0.91 while for the shifts of 40 and 50 mm it remained 0. For the 40 mm shift, the CMS was 38 mm (Velocity), 35 mm (RayDeformable) and 0.8 mm (SmartAdapt) and for the 50 mm shift the CMS was 47 mm (Velocity), 46 mm (RayDeformable) and 1.3 mm (SmartAdapt). The DIR accuracy for the shifts in the LR and AP directions was similar as in IS, with DSC ≥ 0.90 and CMS ≤ 0.9 mm for SmartAdapt, DSC ≥ 0.93 and CMS ≤ 0.1 mm for Velocity and DSC ≥ 0.98 and CMS ≤ 0.1 mm for RayDeformable. The manual delineation of the tumor that was done three times in a row on the un-deformed CT images resulted in DSC ≥ 0.93 and CMS ≤ 0.8 mm when compared to the initial delineation of the structures. The image slice thickness did not have any impact on the results.

In the second part of the phantom study, the tumor was kept static while a change of the tumor volume was introduced between the image series to be registered. For the Velocity and SmartAdapt software solutions this decreased the accuracy of the deformations compared to the results in Fig. 1, while the accuracy of the RayDeformable software was not affected. Fig. 2 shows the DSC for the comparison between the tumor structure in the deformed image and the tumor structure in the target image for ten repetitive deformable registrations. The CMS between the deformed tumor and the tumor in the target image series was ≤ 1.6 mm for all the deformed registrations with the SmartAdapt software, ≤ 1.1 mm with the Velocity software and ≤ 0.1 mm with the RayDeformable software. For the SmartAdapt and Velocity software solutions the tumor shrinkage from 30 or 20 mm to 10 mm resulted in the lowest DSC, while tumor growth was handled better. In the cases of tumor growth, the SmartAdapt software underestimated the volume of the tumor, while in the cases of tumor shrinkage the deformed volume was overestimated. This tendency was not seen for the Velocity or RayDeformable software. The deformation accuracy is not affected by image slice thickness in a general way. Some of the deformable registrations result in a higher DSC when performed on the CT image series with slice thickness of 3 mm and some of the registrations result in a higher DSC when performed on the CT image series with slice thickness of 1 mm.

The combined effect of a change in tumor volume and a shift in position of the tumor in the thorax phantom is presented in Fig. 3. For comparison, the effect of a change in tumor volume without a shift in position is also included. The algorithm in the RayDeformable software solution was the only one that could reproduce the tumor shrinkage from $\phi = 30$ mm to 10 mm but only in the case where the tumor did not move. When this change in tumor volume was combined with a shift, none of the DIR software succeeded. Tumor growth from $\phi = 10$ mm to 30 mm was handled better, with DSC ≥ 0.94 for five of the nine cases. However, none of the DIR software could reproduce the growth in combination with a tumor shift of 30 mm.

3.2. Patient study

For patient cases, the deformations are more complex since multiple structures are deformed at the same time. For the cases included in this study, all the DIR software reproduced the position and shape of the organs at risk (ipsilateral lung, spinal cord, heart) with a DSC ≥ 0.83, with RayDeformable generally obtaining the highest DSC (≥ 0.89), Fig. 4. The position and shape of the GTV could be reproduced by all

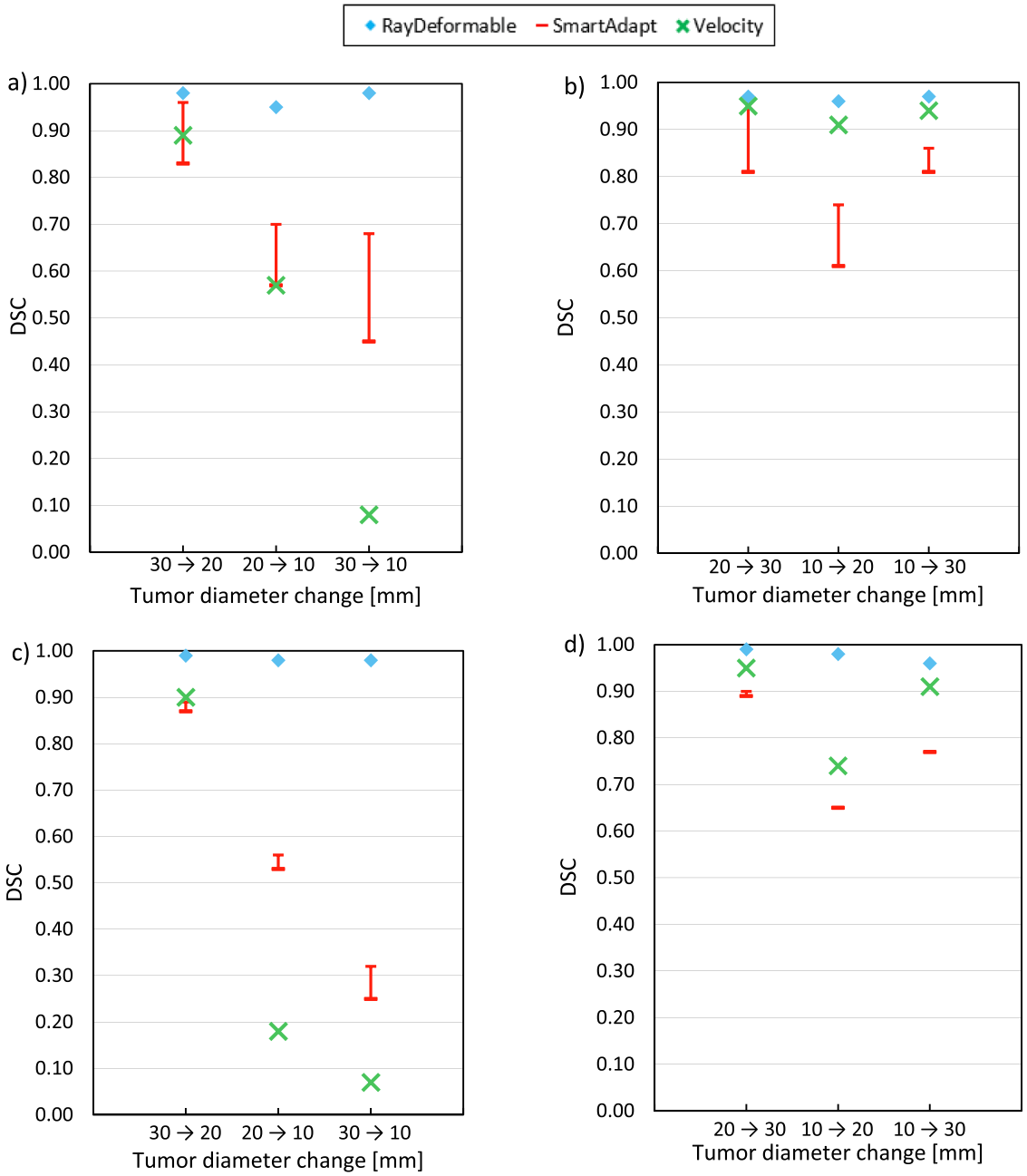

Fig. 2. Dice similarity coefficient (DSC) for the comparison between the tumor structure in the deformed image series and the tumor structure in the target image series for ten repetitive deformations performed on CT image series with slice thickness of 1 mm (a,b) and 3 mm (c,d) for non-moving tumors. Diameter change X → Y indicates the change in tumor diameter in millimeters between the source and the target image, i.e. the tumor diameter changes from X mm in the source image to Y mm in the target image. Figures a and c represent tumor shrinkage and figures b and d represent tumor growth. The obtained DSC varied between deformations with SmartAdapt (indicated as a red interval illustrating the maximum and minimum DSC) while Velocity and RayDeformable gave the same results for each repetition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

software only for two of the patient cases (pat 4 and pat 5), with DSC ≥ 0.88. Both these patients had a GTV shift ≤1.6 cm (see table 1). For patient 2 and 6 with a GTV shift of 2.3 cm and 2.7 cm respectively, the SmartAdapt and RayDeformable software solutions could reproduce the GTV with a DSC ≥ 0.90 while the Velocity software solution could not (DSC ≤ 0.40) even though the stacked VOI-method was used. For patient 1 all the software resulted in DSC = 0 while for patient 3 the highest DSC was 0.68 and was obtained with SmartAdapt. The DSC and CMS for the GTV are presented in table 2 for all the patients. The size of the VOI for the deformed registrations in the SmartAdapt and Velocity software solutions was considered not to have any impact on the deformation results as long as it was large enough to include the entire structures that were being deformed in both the source and the target image; ≤0.01 difference in DSC for SmartAdapt and ≤0.02 difference in DSC for Velocity. When inspecting the deformed images, the lowest DSC for which the structures were considered visually acceptable was 0.88 for the GTV, 0.89 for the ipsilateral lung, 0.87 for the spinal cord and 0.89 for the heart.

The manual delineation of the organs at risk and the GTV that was done three times in a row on the un-deformed CT images resulted in DSC ≥ 0.87 and CMS ≤ 1.0 mm when compared to the initial delineation of the structure.

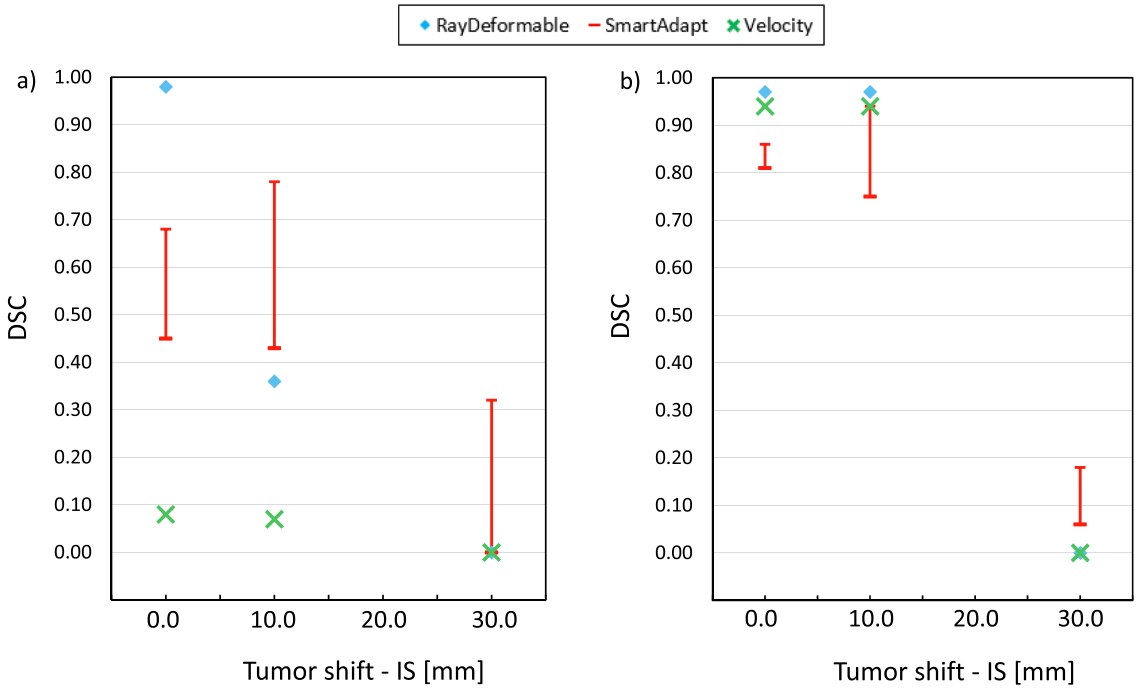

Fig. 3. Dice similarity coefficient (DSC) for tumors with combined shift and volume changes for ten repetitive deformations performed on CT image series with slice thickness of 1 mm. Figure a) shows the results for tumor diameter changes from 30 to 10 mm and b) from 10 to 30 mm. The shift of 0 mm is included for each case for comparison purposes. The obtained DSC varied between deformations with SmartAdapt (indicated as a red interval illustrating the maximum and minimum DSC) while Velocity and RayDeformable gave the same results for each repetition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

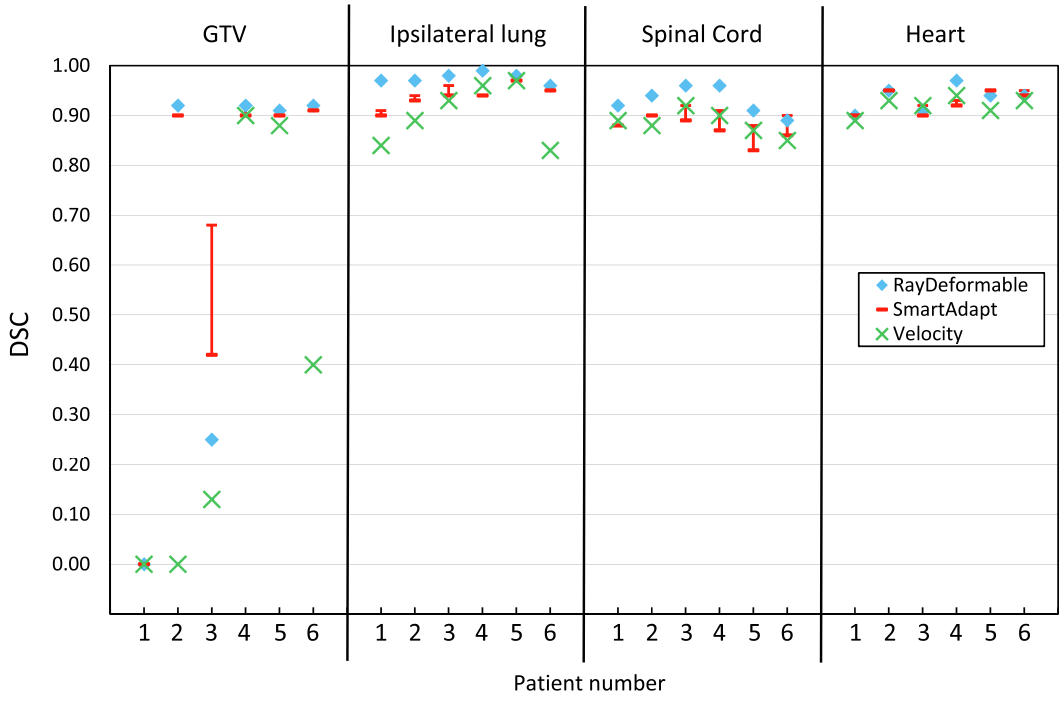

Fig. 4. Dice similarity coefficient (DSC) for the comparison between the structures in the deformed image series and the structures in the target image series for ten repetitive deformations performed on 4DCT image series of patients with moving lung tumors. The obtained DSC varied between deformations with SmartAdapt (indicated as a red interval and illustrating the maximum and minimum DSC) while Velocity and RayDeformable gave the same results for each repetition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
The CMS between the deformed GTV and the GTV in the target image together with the dice similarity coefficients (DSC) calculated with RayDeformable, SmartAdapt and Velocity.

Pat. #	CMS GTV [mm]				DSC GTV			
	RayDeformable	SmartAdapt		Velocity	RayDeformable	SmartAdapt		Velocity
		Max	Min			Max	Min	
1	46	53	45	56	0	0	0	0
2	0	2	0	19	0.92	0.90	0.90	0
3	15	11	7	16	0.25	0.68	0.42	0.13
4	0	1	0	0	0.92	0.92	0.90	0.90
5	0	2	2	2	0.91	0.90	0.90	0.88
6	0	2	1	25	0.92	0.91	0.91	0.40

Table 3
The difference for the evaluated DVH parameters between dose calculations on the deformed patient image series compared to calculations on the target image series are presented for each DIR software. Negative values indicate that the calculated dose or volume was lower in the deformed image series than in the target image series. All plans are normalized to 100% = prescribed dose. A value on white background indicates a visually successful deformation of the structure, while a value on grey background indicates a visually unsuccessful deformation of the structure.

Patient #	Calculation CT	GTV [Diff in % of prescribed dose]		Ipsilateral lung [Diff in % of total structure volume]		Spinal Cord [Diff in % of prescribed dose]	Heart [Diff in % of prescribed dose]
		$D_{98\%}$	D_{Mean}	V_{20Gy}	V_{12Gy}	$D_{2\%}$	$D_{2\%}$
1	RayDeformable	-5.5	8.7	1.0	2.9	0.9	1.2
	SmartAdapt	-9.3	0.7	1.1	3.6	1.3	1.3
	Velocity	0.0	12.7	2.5	4.7	1.3	1.3
2	RayDeformable	0.0	-0.3	-0.3	-0.3	0.2	0.3
	SmartAdapt	0.0	-2.0	-2.7	-3.0	0.7	1.3
	Velocity	-26.7	18.7	-0.3	-0.6	1.3	2.0
3	RayDeformable	-4.2	0.3	-0.1	-0.3	0.0	-0.1
	SmartAdapt	-7.3	-3.3	0.3	1.1	0.0	0.0
	Velocity	-4.7	-1.3	0.1	0.3	0.0	0.0
4	RayDeformable	-0.6	-0.2	-0.2	0.2	0.8	-0.5
	SmartAdapt	-1.4	-1.4	-1.0	-0.8	-1.4	0.0
	Velocity	-1.4	0.0	-0.8	0.0	0.0	1.4
5	RayDeformable	-0.6	0.8	0.8	1.3	0.4	-0.3
	SmartAdapt	-1.3	0.7	0.0	0.0	-0.7	-0.7
	Velocity	-0.7	0.7	0.8	1.2	0.7	0.0
6	RayDeformable	-0.3	-0.3	-0.8	-1.0	-1.2	-1.2
	SmartAdapt	-3.0	0.0	-1.9	-1.9	-1.0	-3.0
	Velocity	-28.0	-4.0	3.4	-3.1	-1.0	0.0

3.3. Dose calculations

When performing dose calculations on the visually successful deformed image series of the patients compared to calculations on the undeformed target image series, the maximum difference in the evaluated DVH parameters was ≤3.0% of the prescribed dose for the GTV, spinal cord, and heart and ≤3.6% of the total structure volume for the ipsilateral lung (Table 3). For the deformed structures that were not considered visually successful, the difference was up to 28% of the prescribed dose and up to 4.7% of the total structure volume. These deformations are marked with a grey background in Table 3.

4. Discussion

The systematic analysis that was carried out on the deformations in the thorax phantom showed that a moving tumor in the phantom could be reproduced with similar uncertainty as in manual delineation by all the DIR software solutions if the tumor did not move > 30 mm or changed its volume. For deformations including volume changes, the results deteriorated down to the level of no overlap at all between the deformed structure and the structure in the target image in some cases. RayDeformable was the software solution that could handle volume changes the best, however as the volume changes were combined with a shift of 30 mm, none of the algorithms succeeded with the deformation. The RayDeformable software solution uses anatomical guidelines which provide information on how specific structures may deform. This may be the reason why volume changes are handled better than when using the SmartAdapt and Velocity software solutions. The difficulties in reproducing large internal shifts (> 30 mm) with the RayDeformable and Velocity software solutions can be due to that these algorithms require the deformed grid to be smooth in the entire image. This means that the voxels in a moving structure (i.e. the tumor or the lungs) cannot move more than a certain distance relative to the surrounding structures without distorting the rest of the image. For the Velocity software

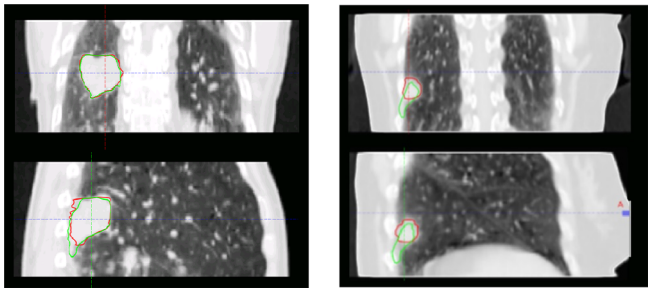

Fig. 5. Frontal (upper figures) and sagittal (lower figures) views of deformed images showing the difficulties of the SmartAdapt software solution to determine the extension of a volume when the volume boundaries were located near voxels with the similar intensites. In both cases the deformed GTV (green) extends into the thorax wall, while the GTV in the target image (red) does not. The DSC for the GTV on the left (patient #5) is 0.90 and for the GTV on the right 0.42 (patient #3). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

solution this is true even if you only deform the image within a volume of interest (VOI) that only covers part of the image. The grid between the VOI and the rest of the image still has to be smooth. In the SmartAdapt software solution this problem is less pronounced since the grid between the VOI in which the DIR is performed and the rest of the image does not have to be smooth. This means that larger shifts can be accomplished without distorting the surroundings if the VOI is set to cover only the moving structure.

One major difference that was observed between the DIR software solutions was that the results obtained with SmartAdapt had a potentially different outcome every time a DIR was performed even though the DIR was repeated with the same settings. This was not seen with Velocity or RayDeformable. This could be a consequence of how the optimization functions for the different algorithms are designed. SmartAdapt is designed to minimize the intensity differences between the deformed image and the target image only, while Velocity and RayDeformable have additional penalization terms. This means that with SmartAdapt the optimal solution for the optimization function may be reached in different ways (i.e. different appearances of the deformed image may result in the same image intensity differences) while for Velocity and RayDeformable the additional penalization terms results in only one optimal solution for each set of DIR input. This means that the result of a DIR in SmartAdapt with low conformity between the deformed structure and the structure in the target image does not exclude the possibility of obtaining a higher conformity if the deformation process is repeated a second time. One tendency that was seen was that the spread of the possible DSC obtained with SmartAdapt often was larger for DIRs performed on the CT series that were collected with an image slice thickness of 1 mm instead of 3 mm. This may be because the series that are collected with 1 mm consist of more images than the series collected with 3 mm. More images means that there are more ways to combine the image appearances to reach the minimum intensity difference between the images. Except this difference there was no clear result suggesting that DIRs performed on CT image series with a slice thickness of 1 mm gave better or worse results than DIRs performed on CT image series with a slice thickness of 3 mm. At our hospital, the most common slice thickness used clinically in reconstructed image series is 2 mm. Therefore this is the slice thickness used for the patient cases in this study. Although, this was not used for the phantom cases which were used to study the influence of image slice thickness. The influence of image slice thickness on the deformation results was suspected to be small and we therefore wanted to investigate the largest clinically relevant difference. In our experience, the slice thickness commonly used clinically can be down to 1 mm but is most often not larger than 3 mm. Therefore, the slice thickness and reconstruction intervals for the study on the impact of the image slice thickness on the deformation results were chosen to be 1 and 3 mm.

For the patient cases the results are more complex since multiple structures change position and volume between the source and the target image series for each DIR. The most difficult structure to reproduce was the GTV which was the only structure with motion exceeding 1 cm. The organs at risk did not move to the same extent and were therefore reproduced in a better way i.e. a higher DSC. According to the publication by the AAPM-TG132 [30] the contouring uncertainty of a structure lies in the DSC-interval of 0.80–0.90. They also point out that DSC calculations are dependent on the volume of the structure, therefore very large or very small structures may have different expected DSC values for contour uncertainty. In this study the lowest observed DSC value for which a contour was considered visually acceptable was 0.87. Of all the recorded DSC in the patient cases only 3/36 values were in the interval 0.80–0.86 (ipsilateral lung with Velocity for pat 1 and 6 and spinal cord with velocity for pat6). These cases were considered visually unacceptable in our study. However, only minor changes in the structure outline would have been necessary to adjust the structures to be visually acceptable. All the software solutions have tools to manually correct the deformed structures if they do not overlap with the target structures in an acceptable way. If the differences are small it can be done without distorting the surrounding tissue, but if larger modifications are needed the surrounding tissue is distorted to unacceptable levels. In this study, no manual corrections have been done for any of the presented results in order to minimize the human impact. Furthermore, both Velocity and RayDeformable have the possibility to use help structures to improve the DIR for the selected structures. However since these options can distort the remaining structures in the image they have not been used in this evaluation.

One limitation that was seen for the deformations that were performed with the SmartAdapt software solution was to correctly determine the boundaries of a structure when it was located close to voxels with similar intensities as the voxels within the structure. This is due to the high weight on the intensity-based approach of the algorithm. Since the objective of the optimization process is to minimize the intensity differences between images, the distinction between a structure and the adjacent surrounding tissue with similar voxel intensity does not affect the mathematical outcome of the optimization since the intensities are similar. Fig. 5 illustrates this problem where the deformed GTV (green) extends into the thorax wall while the GTV in the target image (red) does not. The case on the left shows the deformation of a large tumor (ϕ = 62 mm, pat 5) with a CMS of 16 mm in the IS direction and the case on the right shows a smaller tumor (ϕ = 25 mm, pat 3) with a CMS of 23 mm in IS direction. The DSC for the GTV is 0.90 and 0.42, respectively. For the RayDeformable and Velocity software solutions this issue was not seen in this study, probably due to the anatomical constraints that are used by these algorithms, however similar issues have previously been reported for the Velocity algorithm [32].

For the dose calculations, the effects on the DVH parameters of a non-successful deformation were highly dependent on where within the actual treatment fields the deformed structure (i.e. GTV) was located. Since the treatment plans used in this study were optimized to give the prescribed dose to the entire PTV which was defined based on the ITV using 4DCT information about the GTV motion, the dose to the deformed GTV will not change significantly as long as it remains within the PTV, i.e. within the treatment field. Therefore there is no clear relation between DSC or CMS and the calculated dose deviation to the GTV since these metrics do not reflect if the GTV is still within the high dose region or not. A similar conclusion was drawn by Moriya et al [32] that states that the magnitude of the dose accuracy cannot always be predicted by the magnitude of the DSC value. The number of cases in this study were too limited for a full analysis of the correlation between the dose difference and CMS and DSC. Furthermore such an analysis should not be done on separate dose-volume parameters since this does not give the whole picture of dose deviations in the full dose

distribution. Theoretically the CMS could be a measure that gives information on the probability that the GTV is still located within the high dose region of the field. If the CMS is smaller than the margin between the GTV and the PTV, the probability that the GTV in the deformed image will be located in the high dose region is high. The results of this study show that for all cases with a high DSC (DSC ≥ 0.88) the dose differences are low (≤3% of the prescribed dose) and for lower values of DSC the dose differences are systematically increasing for decreasing DSC for registrations with RayDeformable and Velocity. However, this systematics was not seen for the registrations with SmartAdapt. The take home message from the results of this study is the importance to have a visually acceptable registration. The magnitudes of the reported dose deviations in this study are in accordance with previously reported values [32]. The larger deviations in the DVH parameters were observed in the interface regions between tumor and lung.

In this study the structures have been handled as an entity during the evaluation which provides information on the performance of the algorithms when used for contour propagation or dose estimations due to patient anatomy changes during a treatment course. Future aspects could be to evaluate the performance of the algorithms with regards to the internal voxel shifts for each structure. This would increase the understanding of the algorithms when the intended use is to sum the dose to a moving structure, since each deformed voxel must be in the correct position in order to sum the dose between two or more images correctly.

5. Conclusions

The DIR-algorithms that were evaluated in this study were found to be capable of deforming structures in the thorax region with similar uncertainty as intra-observer delineations when volume changes were not present and shifts were ≤30 mm in phantoms or ≤16 mm in patients. The take home message from the results of this study is the importance to have a visually acceptable registration. To ensure correctness of a DIR, visual inspection of the deformed structures is recommended for every deformation process since the DSC and CMS are not predictive of the associated dose deviation. For a visually acceptable DIR, differences in the evaluated DVH parameters up to 3.0% of the prescribed dose or up to 3.6% of the total structure volume were observed.

Conflict of interest statement

The authors have no conflicts of interest to disclose.

Funding

This study was supported by the Health & Medical Care Committee of the Regional Executive Board, Region Västra Götaland and Futurum – The Academy of Health and Care, Region Jönköping County, Sweden.

References

[1] Machaty M. Higher BED is associated with improved local-regional control and survival for NSCLC treated with chemoradiotherapy: an RTOG analysis. Int J Radiat Oncol Biol Phys 2005;63:40.
[2] McGarry RC, Papiez M, Williams M, Whitford T, Timmerman RD. Stereotactic body radiation therapy of early-stage non-small-cell lung carcinoma: phase I study. Int J Radiat Oncol Biol Phys 2005;63:1010–5.
[3] Wulf J, Baier K, Mueller G, Flentje MP. Dose-response in stereotactic irradiation of lung tumors. Radiother Oncol 2005;77:83–7.
[4] Sarudis S, Karlsson Hauer A, Nyman J, Bäck A. Systematic evaluation of lung tumor motion using four-dimensional computed tompgraphy. ACTA Oncol 2017;Jan;11:1–6.
[5] Keall P, Mageras G, Balter J, et al. The management of respiratory motion in radiation oncology report of AAPM task group 76. Med Phys 2006;33:3874–900.
[6] Miura Hideharu, Ozawa Shuichi, Kawabata Hideo, Doi Yoshiko, Kenjou Masahiro, Furukawa Kengo, et al. Method of evaluating respiratory induced organ motion by vector volume histogram. Phys Med 2016;32(12):1570–4.
[7] Juhler-Nottrup T, Korreman S, Pedersen A, et al. Intrafractional changes in tumor volume and position during entire radiotherpay courses for lung cancer with respiratory gating and image guidance. Acta Oncol 2008;47:1406–13.
[8] Fox J, Ford E, Redmond K, Zhou J, Wong J, Song DY. Quantification of tumor volume changes during radiotherapy for non-small-cell lung cancer. Int J Radiat Oncol Biol Phys 2009;74:341–8.
[9] Knap M, Hoffmann L, Nordsmark M, Vestergaard A. Daily cone-beam computed tomography used to determine tumour shrinkage and localisation in lung cancer patients. Acta Oncol 2010;49:1077–84.
[10] Britton K, Starkshall G, Liu H, et al. Consequences of anatomic changes and respiratory motion on radiation dose distributions in conformal radiotherapy for locally advanced non-small-cell lung cancer. Int J Radiat Oncol Biol Phys 2009;73:94–102.
[11] Lykkegaard Schmidt M, Hoffman L, Kandi M, Moller D, Ruugaard Poulsen P. Dosimetric impact of respiratory motion, interfraction baseline shifts, and anatomical changes in radiotherapy of non-small cell lung cancer. Acta Oncol 2013;52:1490–6.
[12] Court L, Wagar M, Berbeco R, et al. Evaluation of the interplay effect when using RapidArc to treat targets moving in the craniocaudal or right-left direction. Med Phys 2010;37:4–11.
[13] Dowdell S, Grassberger C, Sharp GC, Paganetti H. Interplay effects in proton scanning for lung: a 4D Monte Carlo study assessing the impact of tumor and beam delivery parameters. Phys Med Biol 2013;58:4137–56.
[14] Li X, Yang Y, Li T, Fallon K, Heron D, Huq M. Dosimetric effect of respiratory motion on volumetric-modulated arc therapy–based lung SBRT treatment delivered by TrueBeam machine with flattening filter-free beam. JACPM 2013;14:195–204.
[15] Ong C, Dahele M, Slotman B, Verbakel W. Dosimetric impact of the interplay effect during stereotactic lung radiation therapy delivery using flattening filter-free beams and volumetric modulated arc therapy. Int J Radiat Oncol Biol Phys 2012;86:743–8.
[16] Ong C, Verbakel W, Cuupers J, Slotman B, Senan S. Dosimetric impact of interplay effect on rapidarc lung stereotactic treatment delivery. Int J Radiat Oncol Biol Phys 2011;79:305–11.
[17] Rao M, Wu J, Cao D, Wong T, Mehta V, Shepard D, et al. Dosimetric impact of breathing motion in lung stereotactic body radiation therapy treatment using image-modulated radiotherapy and volumetric modulated arc therapy. Int J Radiat Oncol Biol Phys 2012;83:251–6.
[18] Brock K. Results of a multi-institution deformable registration accuracy study (MIDRAS). Int J Radiat Oncol Biol Phys 2010;76:583–96.
[19] Stanley N, Zhong H, Glide-Hurst C, Chetty I, Movsas B. Systematic evaluation of a deformable image registration algorithm from a commercial software package. Med Phys 2012;39:3876.
[20] Lin H, Ayan A, Zhu T, Both S, Zhai H. A quantitative evaluation of velocity AI deformable image registration. Med Phys 2010;37:3126.
[21] Katsuta, et al. Assessment of a commercially available automatic deformable image registration. IFMBE 2013;39:1849–52.
[22] Noriyuki K, Yukio F, Yoshiyuki K, Suguru D, Ken T, Kazuma K, et al. Evaluation of various deformable image registration algorithms for thoracic images. J Rad Res 2013.
[23] Elstrøm UV, Wysocka BA, Muren LP, Petersen JB, Grau C. Daily kV cone-beam CT and deformable image registration as a method for studying dosimetric consequences of anatomic changes in adaptive IMRT of head and neck cancer. Acta Oncol 2010;49:1101–8.
[24] Kadoya K, et al. Multi-institutional validation study of commercially available deformable image registration software for thoracic images. Int J Rad Onc Biol Phys 2016;96:422–31.
[25] Weistrand O, Svensson S. The ANACONDA algorithm for deformable image registration in radiotherapy. Med Phys 2015;42:40–53.
[26] CIRS-008A PB 111814 dynamic thorax phantom user guide. CIRS Inc. Norfolk, Virginia, USA; 2013.
[27] Fox T, Andl G, Bose S. Velocity and deformable image registration. Varian clinical perspectives, varian medical systems. 2018.
[28] Wang H, Dong L, ÓDaniel J, et al. Validation of an accelerated 'deamons' algorithm for deformable image registration in radiation therapy. Phys Med Biol 2005;50:2887–905.
[29] Thirion JP. Image matching as a diffusion process: an analogy with Maxwell's deamons. Med Image Anal 1998;2:243–60.
[30] Brock K, Mutic S, McNutt T, Li H, Kessler L. Use of image registration and fusion algorithms and techniques in radiotherapy: report of the AAPM Radiation Therapy Committee Task Group No. 132. Med Phys 2017;44:e43–76.
[31] Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26:297–302.
[32] Moriya Shunsuke, Tachibana Hidenobu, Kitamura Nozomi, Sawant Amit, Sato Masanori. Dose warping performance in deformable image registration in lung. Phys Med 2017;37:16–23.