BJR

doi: 10.1259/bjr.20140691

Received: 17 October 2014 | Revised: 12 March 2015 | Accepted: 19 March 2015

Cite this article as:
Roussakis YG, Dehghani H, Green S, Webster GJ. Validation of a dose warping algorithm using clinically realistic scenarios. Br J Radiol 2015;88: 20140691.

FULL PAPER

Validation of a dose warping algorithm using clinically realistic scenarios

[1,2,3]Y G ROUSSAKIS, MSc, [1,2]H DEHGHANI, PhD, [3]S GREEN, PhD and [3]G J WEBSTER, PhD

[1]School of Computer Sciences, University of Birmingham, Edgbaston, Birmingham, UK
[2]Physical Sciences of Imaging in the Biomedical Sciences (PSIBS) Doctoral Training Centre, University of Birmingham, Edgbaston, Birmingham, UK
[3]Hall-Edwards Radiotherapy Research Group, Radiotherapy Physics, Queen Elizabeth Hospital, Birmingham, UK

Address correspondence to: Mr Yiannis George Roussakis
E-mail: *yxr141@bham.ac.uk*

Objective: Dose warping following deformable image registration (DIR) has been proposed for interfractional dose accumulation. Robust evaluation workflows are vital to clinically implement such procedures. This study demonstrates such a workflow and quantifies the accuracy of a commercial DIR algorithm for this purpose under clinically realistic scenarios.
Methods: 12 head and neck (H&N) patient data sets were used for this retrospective study. For each case, four clinically relevant anatomical changes have been manually generated. Dose distributions were then calculated on each artificially deformed image and warped back to the original anatomy following DIR by a commercial algorithm. Spatial registration was evaluated by quantitative comparison of the original and warped structure sets, using conformity index and mean distance to conformity (MDC) metrics. Dosimetric evaluation was performed by quantitative comparison of the dose–volume histograms generated for the calculated and warped dose distributions, which should be identical for the ideal "perfect" registration of mass-conserving deformations.
Results: Spatial registration of the artificially deformed image back to the planning CT was accurate (MDC range of 1–2 voxels or 1.2–2.4 mm). Dosimetric discrepancies introduced by the DIR were low (0.02 ± 0.03 Gy per fraction in clinically relevant dose metrics) with no statistically significant difference found (Wilcoxon test, $0.6 \geq p \geq 0.2$).
Conclusion: The reliability of CT-to-CT DIR-based dose warping and image registration was demonstrated for a commercial algorithm with H&N patient data.
Advances in knowledge: This study demonstrates a workflow for validation of dose warping following DIR that could assist physicists and physicians in quantifying the uncertainties associated with dose accumulation in clinical scenarios.

Modern radiotherapy aims to move towards a personalized treatment for each patient with cancer, requiring reliable predictions of an individual's response to a particular therapy and accurate monitoring of treatment delivery, enabling adaptations to the treatment plan as required. To date, typical radiotherapy practice involves the preparation of a treatment plan based on an initial high resolution CT scan of the anatomy to be treated. However, since the treatment is optimized for the anatomy on planning CT (pCT), any changes in a patient's anatomy during the treatment course itself (which may last for up to 8 weeks) could result in a suboptimal treatment. Currently, to account for interfraction movements, a low-resolution, low-dose CT image [typically cone beam CT (CBCT) or mega-voltage CT (MVCT), although other options exist] of the patient is often acquired prior to each treatment (daily images). This is termed image-guided radiotherapy (IGRT).[1]

In 1997, Yan et al[2] proposed the concept of adaptive radiotherapy (ART), suggesting the adaptation of the treatment plan to account for interfraction anatomical variations, based on these daily images. Such treatment adaptations are sometimes currently employed in routine clinical practice when significant anatomical changes are observed, such as substantial weight loss.[3] State-of-the-art ART, on the other hand, aims to regularly monitor the treatment delivery and adapt when necessary (offline ART)[2] or even predict the result and alter it before the treatment of that day (online ART).[4] The ability to determine the accumulated delivered dose to deforming anatomy is of vital importance not only for ART but also for the assessment and optimization of radiobiological models,[5] since without it, these models are informed by less accurate estimates of delivered dose to each tissue or partial tissue volume. However, certain limitations such as inaccuracies in contour propagation and in reliable dose

accumulation currently prevent efficient routine monitoring of delivered dose throughout the treatment.

Deformable image registration (DIR) algorithms have been proposed as a method for facilitating these processes. The accuracy of DIR algorithms is therefore of critical importance and has been the subject of investigation by several researchers, with mechanical phantoms,[6–13] patient images[14–22] and digital phantoms (*i.e.* patient images artificially deformed with known deformations)[10,11,23] being extensively used for DIR assessment.

An extension to these issues is the application of the underlying anatomical deformations to a calculated dose distribution, which is a necessary step in interfractional dose accumulation. Such dose warping process involves the direct deformation of a calculated dose distribution by applying the deformation matrix estimated during DIR between two anatomical scans, essentially warping the dose according to the reference anatomy. Dose warping and deformable dose accumulation have been employed in a number of clinical investigations, including a dose feedback technique in ART frameworks,[24] the assessment of planning target volume (PTV) margins[25] and the examination of parotid gland dose–effect relationships,[26] based on dose distributions recalculated on daily or weekly scans and the accumulation on a single frame of reference. Consequently, quality assurance and evaluation techniques have been investigated in order to validate the applicability of this dose warping concept. Previous work has investigated mathematical models to directly convert DIR errors into dose-warping uncertainties, through the use of patient images and mechanical or digital phantoms,[15,27–30] while a number of deformable dosimetric and non-dosimetric gel phantoms have been produced enabling the experimental evaluation of both DIR and dose warping.[31–35] Even though some of these studies revealed promising results, they have not convinced the radiotherapy community that these uncertainties are adequately understood.[36]

In one such study, Yeo et al[34] used a cylindrical deformable dosimetric gel phantom for the experimental validation of dose warping against actual three-dimensional (3D) measurements. The warped and measured dose distributions revealed an agreement of 3D $\gamma_{3\%/3mm} = 99.9\%$, after small deformations (approximately 9 mm), and $\gamma_{3\%/3\,mm} = 96.7\%$ after larger deformations (approximately 20 mm). The authors therefore concluded that "dose-warping may be justified for small deformations in particular and those that do not involve significant density changes". On the other hand, Juang et al[35] exposed "substantial errors in a commercial DIR" used for dose-warping evaluation, utilizing another 3D deformable dosimetric gel, revealing a 3D $\gamma_{3\%/3mm}$ passing rate of 60%.

Such studies, and especially the use of deformable dosemeters for the evaluation of dose warping, are very important as they can reveal the 3D dosimetric impact owing to the uncertainties of a given DIR algorithm. However, they possess three important limitations: first, typical physical dosimetric phantoms present limited image complexity and would not assess the performance limits of the DIR algorithm under evaluation in clinical scenarios. Second, plan delivery, intrinsic dosimetric and dose reading uncertainties are present when using any type of dosemeter in physical phantom measurements. The third limitation is the fact that even where such approaches can offer high precision dosimetric uncertainty evaluation, they cannot directly inform users about the potential extent of those uncertainties in practical clinical cases. All these issues will be addressed in this work.

In the present study, a workflow for the robust validation of DIR and dose warping is presented, using patient images artificially deformed with clinically realistic deformations and clinically optimized Monte Carlo dose calculations of intensity-modulated radiotherapy (IMRT) plans, quantifying both the spatial errors in the deformable registration and their dosimetric impact when applied to dose accumulation. In contrast to previously proposed evaluation procedures, this method examines and reports dose-warping uncertainties under clinically relevant scenarios. Although the validation workflow is applicable for different DIRs and clinical indications, it is here employed specifically for the evaluation of a commercial software (OnQ rts®; Oncology Systems Limited, Shropshire, UK) in head and neck (H&N) cancer patient cases.

METHODS AND MATERIALS

Data selection

A total of 12 H&N patient data sets, consisting of a digital imaging and communications in medicine CT data set, associated structure set (RTS), 6-MV IMRT plan (RTP) and corresponding dose distribution (RTD), were randomly selected for this retrospective simulation study. Of the 12 patients used, 4 were treated for unilateral and 8 for bilateral H&N cancer, while all treatment plans were created using the Monte Carlo dose calculation algorithm in Monaco® v. 3.20 (Elekta AB, Stockholm, Sweden) treatment planning system (TPS). Treatment plans ranged in complexity with 1–3 target volumes (1 PTV: $n = 5$; 2 PTVs: $n = 3$; 3 PTVs: $n = 4$) while the prescribed dose per fraction ranged from 2 to 3 Gy.

Artificial deformations

For each patient, the pCT data set and structures (pRTS) were transferred to ImSimQA v. 3.0.77 (Oncology Systems Limited, Shropshire, UK), where four clinically realistic artificial deformations were manually introduced to create three "CT*n*" and "RTS*n*" data sets (*i.e.* *n* referring to the *n*th artificial deformation in an assumed interfractional dose accumulation workflow) in a process previously demonstrated by Varadhan et al.[23] ImSimQA employs a radial basis function with thin-plate spline[37] kernel function for the application of global deformations, while to compact support radial basis function[38] is utilized for local deformations.

Backward (Def1) and forward (Def2) neck flexion, weight loss (Def3) and upward shoulder movement (Def4) have been applied to each data set using "global" or "local" deformations, as summarized in Table 1 and shown in Figure 1. Artificial deformations have been based on actual clinical observations during image guidance at our institution (Queen Elizabeth Hospital, Birmingham, UK) and visually inspected by a specialist consultant and a specialist radiographer for clinical relevance. Volume conservation in critical structures was quantitatively investigated by comparing the original and deformed volumes.

Table 1. The four types of clinically realistic artificial deformations applied to the planning CT images within ImSimQA v3.0.77 (Oncology Systems Limited, Shropshire, UK) to create the four "CT*n*"

Name	Description	Details
Def1	Forward neck flexion	Chin and back of head moved by 10–15 mm in opposite directions ("global" deformation)
Def2	Backward neck flexion	Chin and back of head moved by 10–15 mm in opposite directions ("global" deformation)
Def3	Weight loss	Neck region shrank by 5 mm ("local" deformation)
Def4	Upward shoulder movement	Shoulders displaced by 10–15 mm ("local" deformation)

Def1–4, the four artificial deformations applied.

Deformable image registration and dose-warping validation

The performance of the DIR algorithm has been evaluated with the workflow shown in Figure 2. Following the application of clinically realistic artificial deformations on the pCT images, the new deformed images (CT*n*) were sent to Monaco TPS where the original treatment plan was applied with identical conditions (*i.e.* beam arrangement, isocentre, segment positions, monitor units and monitor units per segment). The new dose calculated (Dose_True) was considered the "true" dose, since this would be the distribution delivered to the patient if the original plan was to be applied on this anatomy. CT*n* with the associated structure set (RTS*n*) and the calculated dose distribution (Dose_True) were then loaded to OnQ rts together with pCT, where rigid followed by DIR was performed. pCT was treated as the "reference", and each CT*n* as the "moving" image (*i.e.* deforming back to the original anatomy). OnQ utilizes an intensity-based Demons algorithm[39] for the execution of DIR between two CT data sets. The CT*n*, RTS*n* and associated Dose_True were deformed accordingly by applying the deformation matrix calculated during DIR (dynamic vector field). This resulted in warped image (dCT*n*), structure set (dRTS*n*) and dose distribution (Dose_Warp), which were then also copied to pCT.

A "perfect" DIR algorithm would be able to bring the artificially deformed images and structures back to their original configuration. Any deviations in these objects, therefore, would be owing to the spatial inaccuracies of the DIR algorithm being investigated. Comparison between the original (pRTS) and warped (dRTS*n*) structure sets has been performed and the registration result quantitatively evaluated utilizing the conformity index (CI) and mean distance to conformity (MDC) metrics:

CI is defined as

$$\mathrm{CI}=\frac{V_n \cap V_p}{V_n \cup V_p} \qquad (1)$$

where V_n is the nth warped volume, and V_p is the original planning volume of a certain structure.[40] MDC is defined as "the average distance that all outlying points in the warped volume, V_n, must be moved in order to achieve perfect conformity—overlap—with the original volume, V_p,[40] and is measured in units of distance (*i.e.* millimetre).

Furthermore, for the above situations in which mass of organs under investigation is conserved, the same hypothetical "perfect" DIR algorithm would result in agreement between the dose–volume histogram (DVH) analyses of [Dose_Warp, pRTS] and [Dose_True, dRTS*n*]. Differences in these values can therefore be attributed to DIR inaccuracies and are evaluated by both visual comparison of the DVHs and by quantitative differences in clinically relevant dose metrics. The two dose distributions were also compared with the original plan (Dose_Original) in order to expose the dose delivery errors owing to the introduction of artificial deformation on the original anatomy and the ability of dose warping to estimate this effect. Evaluation was herein performed utilizing the DVHs of brain, brainstem, larynx, spinal cord, contralateral parotid and the clinical target volume (CTV). Spatial evaluation was performed by dose subtraction and 3D gamma analysis [computational environment for radiotherapy research (CERR); Washington University, St Louis, MO] with 2%/1 mm criteria calculating the passing ratio for all voxels receiving >20% of the maximum dose, for Dose_Original, Dose_True and Dose_Warp.

Statistical analysis

The non-parametric two-sided Wilcoxon signed-rank test was used for statistical analysis of dosimetric results, comparing the mean absolute differences in mean, median, minimum or maximum dose within certain structures as calculated by the "Dose_True", "Dose_Warp" and "Dose_Original". Statistical analysis was performed using R programming language v. 3.0.1 (R Core Team, Vienna, Austria; www.r-project.org) and was carried out for each type of artificial deformation individually in order to retain statistical independence.

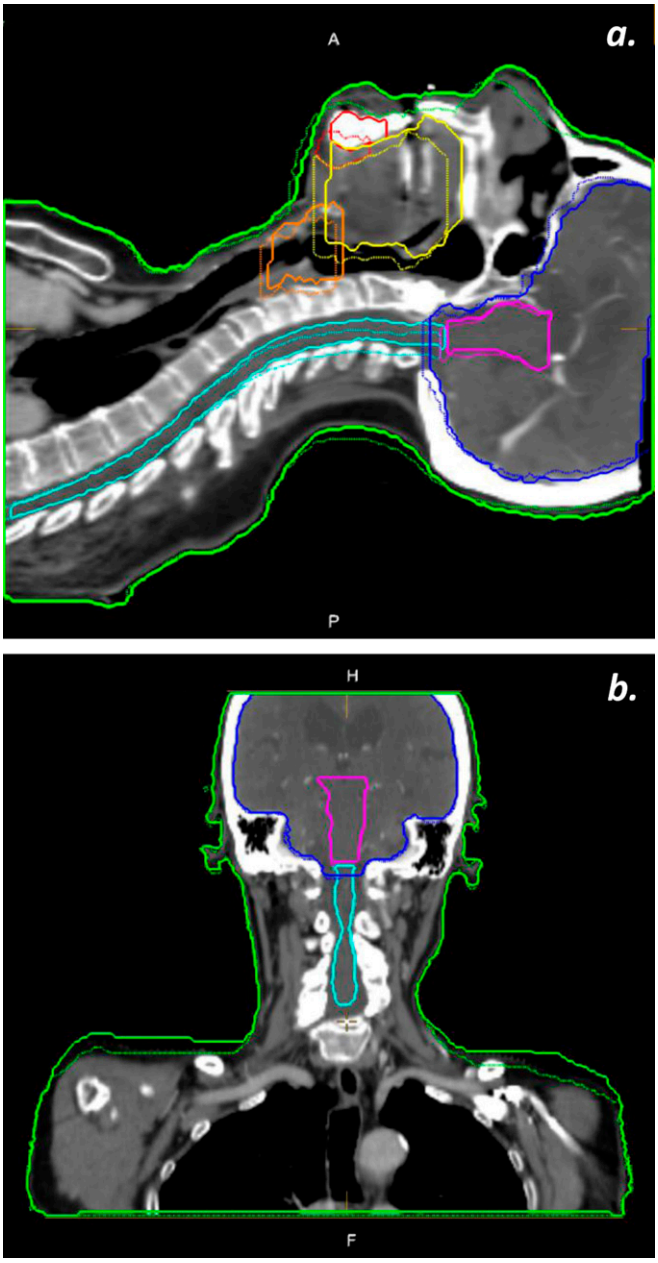
RESULTS

Deformable image registration evaluation

The evaluation of the DIR algorithm's performance under the four artificial test conditions is shown in Figure 3. Figure 3a exposes the respective mean values of MDC over the 12 patient data sets used. With the exception of chiasm, the analysis of all volumes revealed an MDC of 1.2–2.0 mm. Considering the voxel size of the CT scans used for this study (1.2 mm in the x and y directions, and 2.0 mm in the z-direction), it is observed that the average registration result is accurate to within 1–2 voxels.

Figure 3b shows the mean CI value for each artificial deformation for the data sets employed. Brainstem, contralateral

Figure 1. Examples of artificial deformations applied, with dotted lines representing original contours, whereas normal lines are showing new contours: (a) backward neck flexion and (b) upward shoulder movement.



parotid and mandible revealed mean CI values of 0.7 while cord, brain, body and CTV had a CI of ≥0.8. Note that even though CTV is not an anatomically definable structure, it was incorporated in the DIR evaluation process as it would be used for the dosimetric analysis. Chiasm was the organ that revealed the lowest CI values (0.4–0.2). As chiasm only covers 2–3 slices, an average inaccuracy of 1–2 voxels in the *z*-direction results in very low CI.

Dose-warping validation

Examples of single fraction DVHs for the Dose_Original, Dose_True and Dose_Warp for a typical patient in the four artificial deformations are shown in Figure 4. The differences observed between the Dose_Original and Dose_True curves clearly demonstrate the expected dose delivery errors in the presence of the artificial clinically realistic deformations. The warped dose distribution revealed a generally good agreement with the Dose_True. However, regions receiving low dose (*i.e.* <20% of prescribed dose) were occasionally underestimated, as observed in brain, brainstem and spinal cord DVHs in Figure 4.

Figure 5 shows a comparison of the actual dose delivery change owing to the presence of deformations (*i.e.* |Dose_True – Dose_Original|), against the estimated change after warping back to the original anatomy (*i.e.* |Dose_Warp – Dose_Original|), which is subject to errors owing to DIR uncertainties. For a perfect DIR algorithm and when mass is conserved in the regions under investigation, these values would be identical in all cases. This comparison revealed good agreement (0.02 ± 0.03 Gy) with no statistically significant differences ($0.6 \geq p \geq 0.2$ in all cases). It is observed that the minimum dose received by the CTV is occasionally, but not significantly ($0.4 \geq p \geq 0.3$ in all cases), underestimated after dose warping as small spatial uncertainties have bigger dosimetric effects in regions with steep dose gradients.

To further quantify the clinically relevant accuracy of the dose warping process, Dose_Warp was compared against Dose_True in terms of median, mean, maximum or minimum dose to the brain, brainstem, spinal cord, contralateral parotid and CTV, with the observed differences revealing no statistical significance (*i.e.* $0.5 \geq p \geq 0.2$ in all cases).

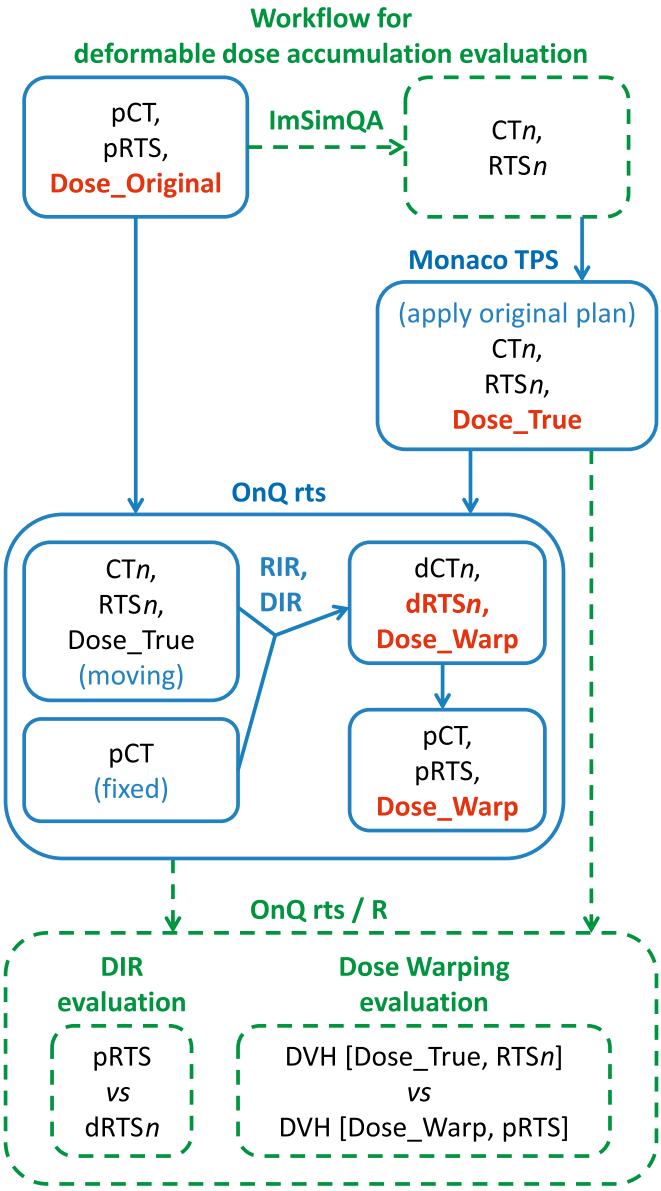
Even though non-statistically significant throughout a total of 48 cases investigated, a number of substantial DIR and dose-warping errors have been observed, as, for example, in Figure 4a; in this instance, Dose_Warp reveals good agreement with Dose_True for all organs except the brainstem, for which a difference of 0.35 and 0.25 Gy in the median and maximum dose estimation was observed, respectively, being the result of an outlier MDC error of 3.2 mm in DIR.

Figure 6 shows examples of gamma maps between Dose_Original, Dose_True and Dose_Warp. Gamma analysis of Dose_Original *vs* Dose_True (Figure 6a) illustrates the differences in dose distribution owing to the applied anatomical deformation. Dose_Original *vs* Dose_Warp (Figure 6b) shows the same differences after warping of Dose_True back to the original anatomy, while Dose_True *vs* Dose_Warp (Figure 6c) illustrates the effect of warping the recalculated dose distribution to the reference anatomy. Subsequent review of the 3D gamma maps confirmed that the regions of largest disagreement are situated in regions that combine dose gradient and displacement, as would be expected, also demonstrated in the dose difference map (Figure 6d). Gamma analysis of Dose_True *vs* Dose_Warp in the forward neck flexion simulation (Figure 6c) revealed greater discrepancies at the chin and neck area, which experienced the greatest displacement, while small differences were observed at the inner body region where anatomical displacement was smaller.

Evaluation of artificial deformations

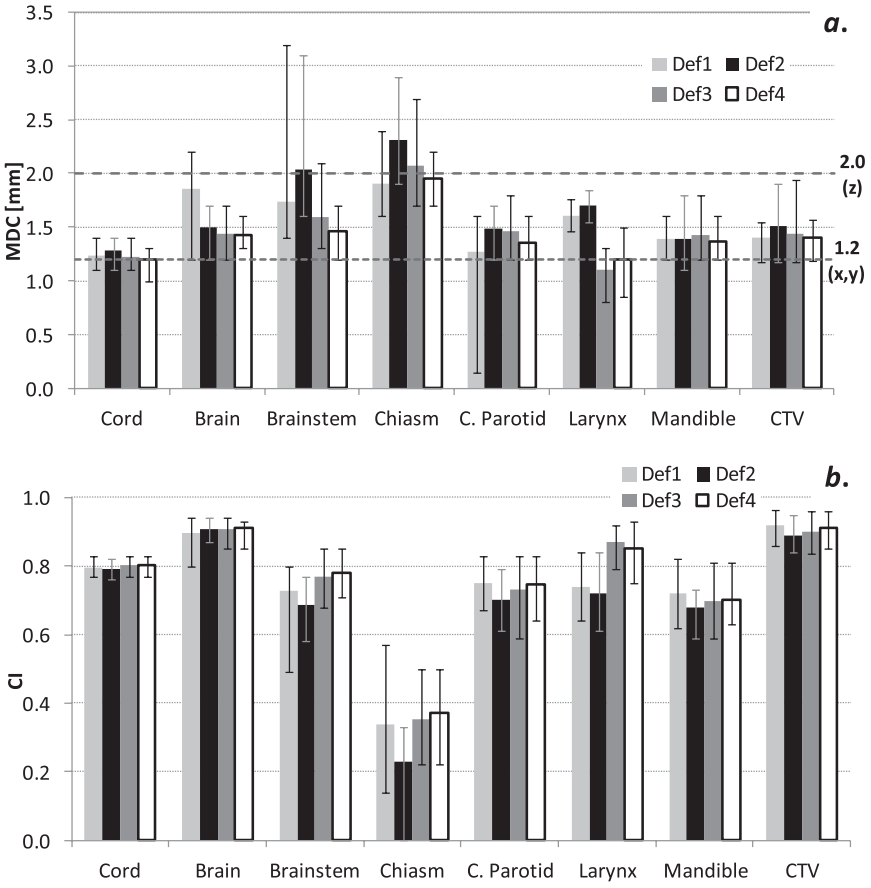
Artificial deformations have been confirmed as clinically realistic after visual inspection by a specialist consultant and

Figure 2. Flow chart summarizing the dose-warping evaluation workflow. A planning CT (pCT) and the associated structure set (pRTS) were imported in ImSimQA software v. 3.0.77 (Oncology Systems Limited, Shropshire, UK), where clinically realistic deformations were applied to create four different artificial CT (CT*n*) and structure sets (RTS*n*)—that is, *n* refers to the *n*th deformation. CT*n* and RTS*n* are then sent to Monaco treatment planning system where the original plan was applied and new dose calculated (Dose_True). Rigid image registration (RIR) followed by deformable image registration (DIR) in OnQ rts® software (Oncology Systems Limited) then warped CT*n* to pCT and the calculated deformation matrix was subsequently applied to RTS*n* and Dose_True, resulting in a warped data set (dCT*n*, dRTS*n* and Dose_Warp). The pRTS and dRTS*n* are quantitatively compared (conformity index and mean distance to conformity) for the evaluation of DIR. Then, dose–volume histograms (DVHs) are created from Dose_True with RTS*n* (*i.e.* actual delivered dose) and Dose_Warp with pRTS (*i.e.* actual delivered dose warped to the reference anatomy). DVHs were compared qualitatively (visual inspection of dose distributions) and quantitatively (comparison of certain measures, including mean, median, maximum dose to each organ) for the evaluation of deformable dose accumulation (DDA). TPS, treatment planning system.



a specialist radiographer. Given the artificial deformations being applied in this work (*i.e.* neck flexion, weight loss and shoulder movement), we would not expect structures such as the brainstem, parotids and perhaps spinal cord to be volumetrically changed. Reassuringly, quantitative comparison of the original and deformed structures revealed perfect agreement in upward shoulder movement and weight loss simulation cases and near perfect volumetric agreement (≥99%) in the forward/backward neck flexion simulation cases, for all structures under examination. An investigation of the impact of this on the DVH analysis employed herein is shown in the Appendix A. The magnitude of the potential errors would not notably impact on the above results.

Figure 3. Evaluation analysis of deformable image registration (DIR) in the 12 patient cases under investigation, for the 4 artificial deformations applied (Def1–4). (a) Average mean distance to conformity (MDC); (b) average conformity index (CI), for spinal cord, brain, brainstem, chiasm, contralateral parotid (C. Parotid), larynx, mandible and clinical target volume (CTV). The error bars in both plots represent the range of values observed, while the horizontal dashed lines in (a) represent the voxel size of the CT scans used in *x*, *y* and *z* direction.



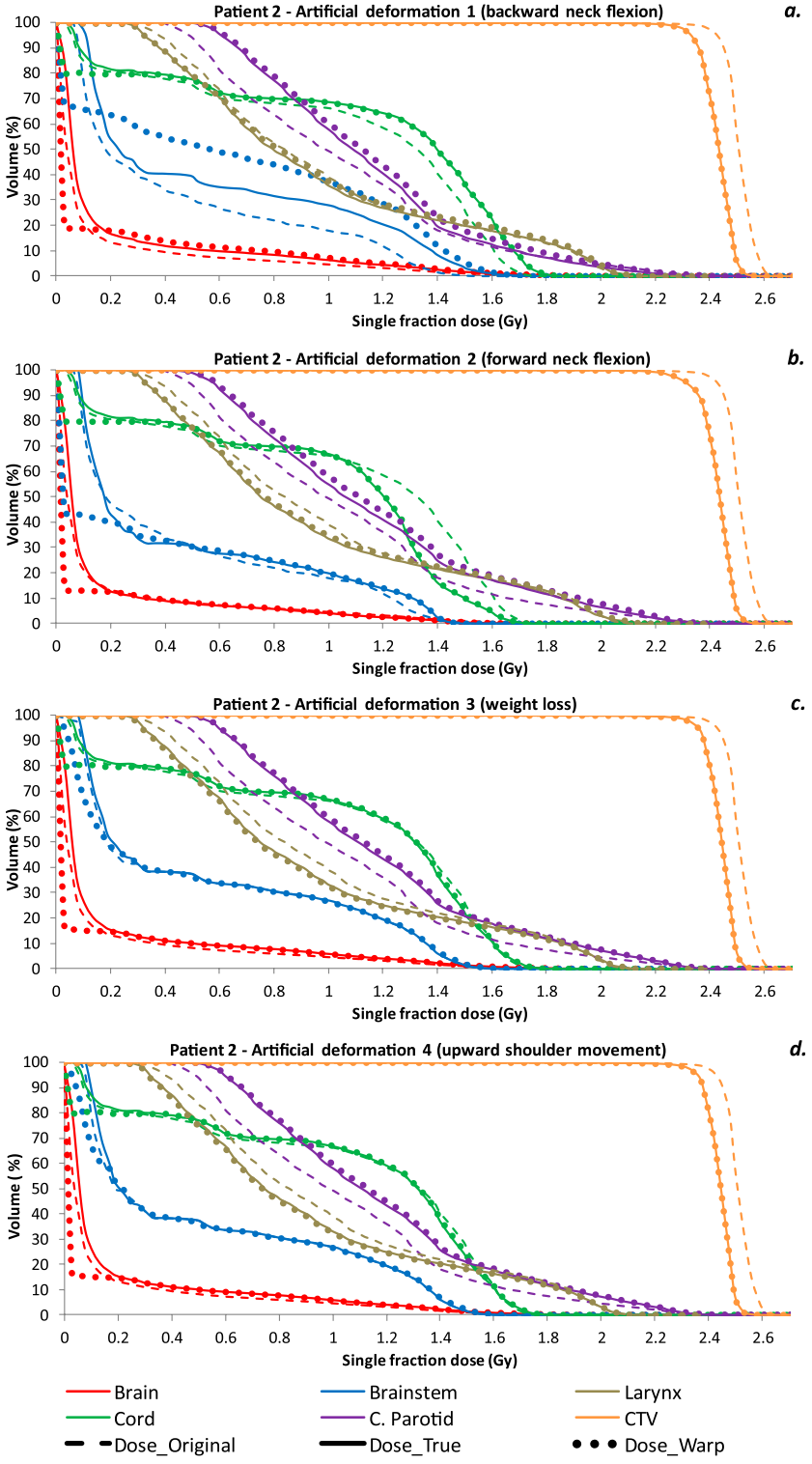
DISCUSSION

This study demonstrates a DIR algorithm validation workflow for image registration and dose warping throughout fractionated radiotherapy, while overcoming the four limitations of recent studies that employed physical phantoms for the evaluation of dose warping[31–35] that were noted earlier: limited image complexity, dose measurement accuracy and transfer of findings to the clinical scenarios. The workflow has been used for the validation of a commercial algorithm which, largely, demonstrated accurate predictions of the actual dose distributions under four clinically realistic deformation scenarios. All analysis was performed for single fraction situations in order to simulate the scenario of daily treatment monitoring that would be most sensitive to any errors, as it excludes averaging effects from daily anatomical variations. Recalculated dose distributions were successfully warped to the reference anatomy [Dose_Warp, pRTS] and revealed good agreement to the ground truth [Dose_True, dRTS*n*], with the observed differences having no statistical significance (Wilcoxon test, $0.5 \geq p \geq 0.2$). However, considerable registration and dose-warping errors have been observed in a small number of cases, a finding that illustrates the importance of such validation work as a means of highlighting and understanding the presence and extent of errors in dose accumulation processes and the need for visual inspection of DIR results. These registration and dose-warping errors were primarily observed in low contrast regions where DIR algorithms are known to have inferior performance.

As the employed workflow included the comparison of DVH metrics for the original and artificially deformed anatomies, any volumetric differences between the structures could potentially result in uncertainties. Volumetric comparison of the original and deformed structures revealed ≥99% agreement, in the forward/backward neck flexion simulation cases. The potential "worst-case scenario" dosimetric errors induced by this small mismatch in the examined test cases were quantified as detailed in the Appendix A. These "worst-case scenario" errors were shown to have very small dosimetric consequences (*e.g.* ±0.005 Gy to the estimation of median dose to the brainstem) and have been ignored.

There are three limitations to the present work. First, deformable registration and dose warping between two pCT quality scans, as herein, is perhaps more robust than would be encountered clinically; even though some radiotherapy centres use CT-on-rails for daily imaging, typical IGRT procedures employ a range of alternative modalities, including CBCT, MVCT or mega-voltage CBCT. These scans have lower image quality and smaller field-of-view than does CT, which may further

Figure 4. Dose–volume histograms comparing the clinically prescribed [Dose_Original, pRTS], recalculated on the artificial image [Dose_True, dRTS*n*], and warped [Dose_Warp, pRTS] dose distributions, of artificial deformations 1–4 (a–d) for a single typical patient. In situations where the mass of these organs is conserved, a "perfect" deformable image registration would show agreement between [Dose_True, dRTS*n*] and [Dose_Warp, pRTS]. C. Parotid, contralateral parotid; CTV, clinical target volume.



compromise DIR performance. The validation workflow performed herein can in principle be applied to artificially deformed images with added noise for the simulation of these situations, an approach that is currently under investigation, while the results of the present study could represent a "best-case" scenario for the use of this algorithm in H&N cases. It

Figure 5. The mean actual dose delivery change introduced by anatomical deformations, |Dose_True – Dose_Original|, against the mean estimated dose change by deformable image registration-based dose warping |Dose_Warp – Dose_Original|, for the 12 patient cases investigated and the 4 artificial deformations applied (a–d), in spinal cord, brain, brainstem, contralateral parotid (C. Parotid), larynx and clinical target volume (CTV). In situations where the mass of these organs is conserved, a "perfect" deformable image registration would result in these values being the same for all situations. Max, maximum; Min, minimum.

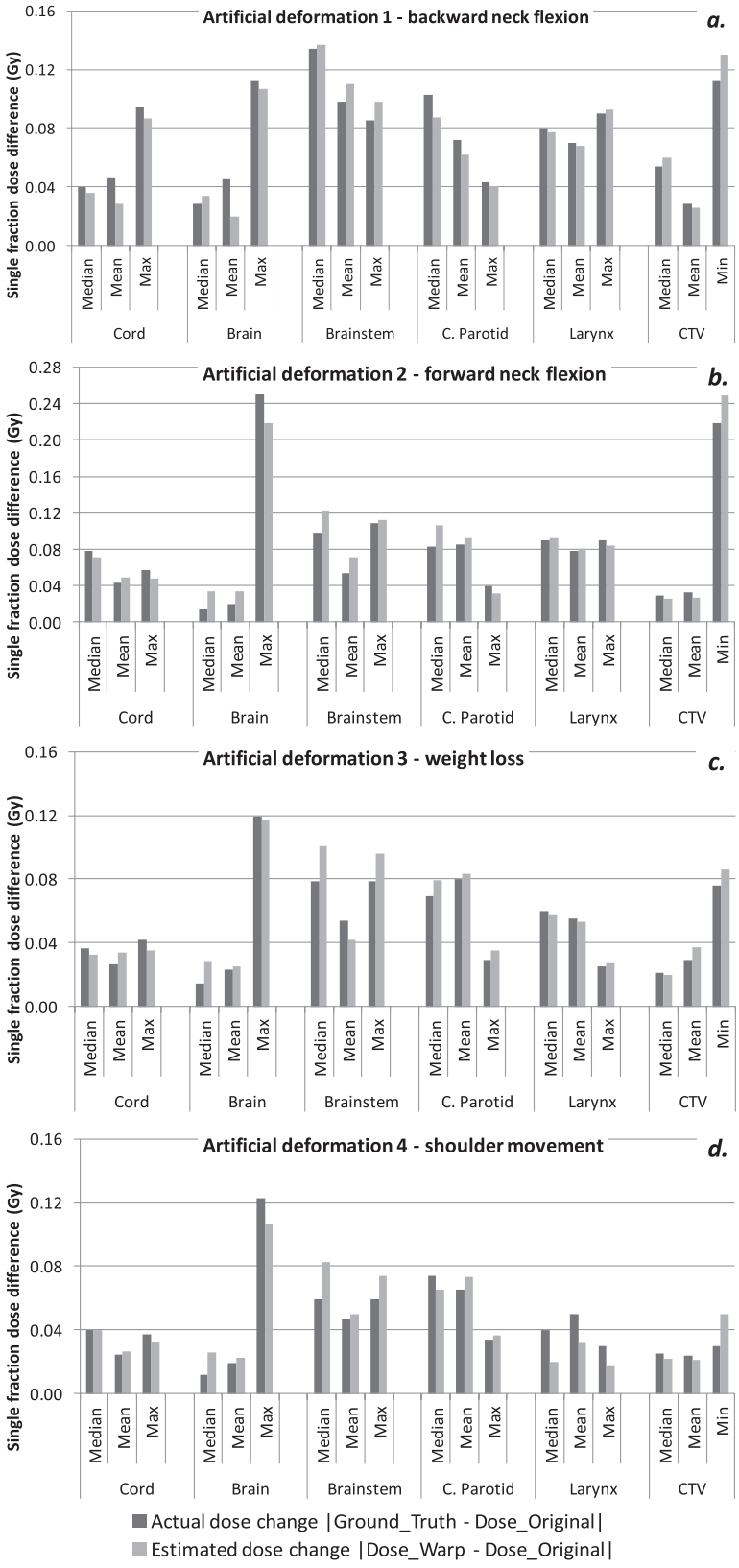
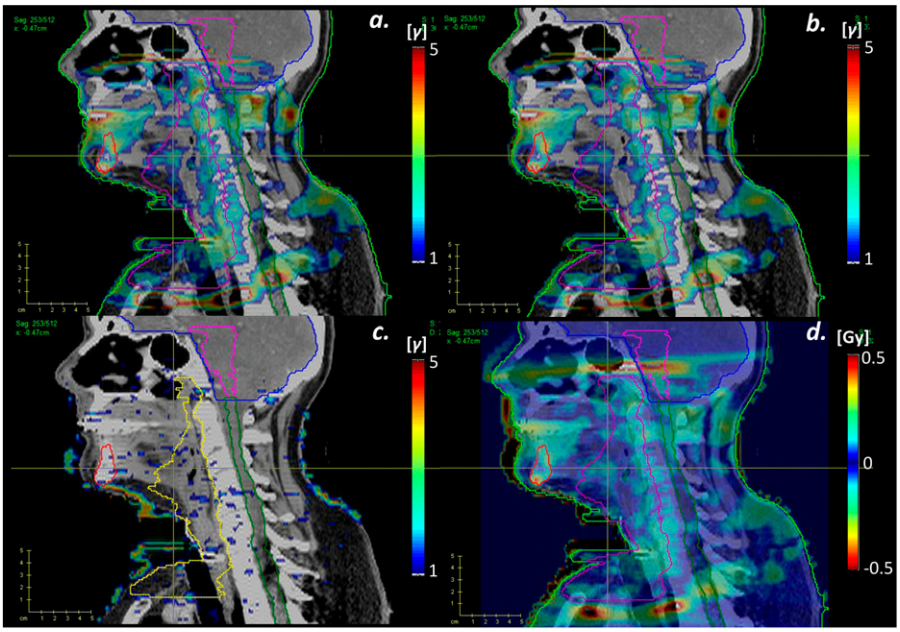


Figure 6. Three-dimensional gamma analysis (2%/1mm criteria) of (a) Dose_Original *vs* Dose_True, (b) Dose_Original *vs* Dose_Warp, (c) Dose_True *vs* Dose_Warp and (d) dose subtraction Dose_Original – Dose_True, of a representative example after forward neck flexion simulation.



should also be emphasized that the same workflow is applicable to validate image registration and dose warping for any anatomical site and any DIR algorithm, provided that the examined structures undergo mass-conserving deformations.

Second, as the proposed workflow employs DVH analysis, the mass and volume of organs under investigation must be conserved. As previously observed, the volume of tumours and certain organs occasionally decreases during the course of fractionated H&N radiotherapy,[41] a situation that was not simulated herein. Nevertheless, this limitation is also present in other dose warping evaluation workflows, such as the ones employing deformable dosimetric gel phantoms, while mass conservation is an underlying assumption in many DIR algorithms.

The third limitation of the present work is the difficulty in quantitatively determining the spatial position of observed differences using, for instance, gamma analysis as an evaluation test, since the warped (Dose_Warp) and calculated (Dose_True) dose distributions are associated with different anatomies. A gamma analysis between them includes effects of both real anatomical change and errors in deformable registration, the separation of which is challenging. DVH analysis is applicable however, which is one of the main examination tools used by physicians and physicists in a clinical setting. This could be construed as an advantage of the current evaluation workflow, providing clinically meaningful and organ-specific uncertainty measures, a feature absent from most validation work in the area. Besides, radiobiological analysis can be applied to further assess the impact of DIR and dose-warping uncertainties, which is beyond the scope of the present study.

As discussed previously, a number of studies have investigated techniques for the evaluation of dose warping using deformable dosimetric and non-dosimetric gel phantoms.[31–35] Even though such evaluation methodologies offer valuable advantages, such as the ability to perform quantitative spatial dosimetric evaluation, their ability to offer comprehensive conclusions under clinical conditions is hindered by limited anatomical complexity, dosimetric and dose-reading uncertainties and inability of clinical interpretation of results. These techniques could therefore be employed as an initial dose warping evaluation workflow, while the procedure described herein could provide further insight and interpretation of results under clinical scenarios.

CONCLUSIONS

This retrospective simulation study demonstrates a workflow for the validation of DIR and dose-warping performance of any DIR algorithm in cases with mass-conserving deformations. Using this workflow, with H&N patient images artificially deformed with clinically realistic deformations, we have confirmed that OnQ rts successfully propagated the actual delivered dose to the original planned anatomy by dose warping following CT-to-CT DIR. Larger errors were occasionally observed, however, confirming that DIR performance should always be evaluated and approved before proceeding to dose warping and accumulation in the clinical setting.

ACKNOWLEDGMENTS

Radiotherapy Physics department at University Hospitals Birmingham has a non-financial research agreement with Oncology Systems Limited.

FUNDING

This work was supported by Engineering and Physical Sciences Research Council grant EP/F50053X/1 funding Physical Sciences of Imaging in Bio-medical Sciences Doctoral Training Centre.

REFERENCES

1. National Cancer Action Team. *National radiotherapy implementation group report. Image guided radiotherapy (IGRT) guidance for implementation and use*. London, UK: NCAT; 2012.
2. Yan D, Wong J, Vicini F, Michalski J, Pan C, Frazier A, et al. Adaptive modification of treatment planning to minimize the deleterious effects of treatment setup errors. *Int J Radiat Oncol Biol Phys* 1997; **38**: 197–206. doi: 10.1016/s0360-3016(97)00229-0
3. Shrimali RK, Webster GJ, Lee LW, Bayman N, Sheikh HY, Stratford J, et al. Reactive plan adaptation for lung radiotherapy based on cone beam CT—the Christie experience. *Clin Oncol* 2011; **23**: S17. doi: 10.1016/j.clon.2011.01.356
4. Court LE, Tishler RB, Petit J, Cormack R, Chin L. Automatic online adaptive radiation therapy techniques for targets with significant shape change: a feasibility study. *Phys Med Biol* 2006; **51**: 2493–501. doi: 10.1088/0031-9155/51/10/009
5. Jaffray DA, Lindsay PE, Brock KK, Deasy JO, Tomé WA. Accurate accumulation of dose for improved understanding of radiation effects in normal tissue. *Int J Radiat Oncol Biol Phys* 2010; **76**(Suppl. 3): S135–9. doi: 10.1016/j.ijrobp.2009.06.093
6. Kirby N, Chuang C, Ueda U, Pouliot J. The need for application-based adaptation of deformable image registration. *Med Phys* 2013; **40**: 011702. doi: 10.1118/1.4769114
7. Kirby N, Chuang C, Pouliot J. A two-dimensional deformable phantom for quantitatively verifying deformation algorithms. *Med Phys* 2011; **38**: 4583–6. doi: 10.1118/1.3597881
8. Kashani R, Hub M, Balter JM, Kessler ML, Dong L, Zhang L, et al. Objective assessment of deformable image registration in radiotherapy: a multi-institution study. *Med Phys* 2008; **35**: 5944–53. doi: 10.1118/1.3013563
9. Bender ET, Tomé WA. The utilization of consistency metrics for error analysis in deformable image registration. *Phys Med Biol* 2009; **54**: 5561–77. doi: 10.1088/0031-9155/54/18/014
10. Wang H, Dong L, O'Daniel J, Mohan R, Garden AS, Ang KK, et al. Validation of an accelerated "demons" algorithm for deformable image registration in radiation therapy. *Phys Med Biol* 2005; **50**: 2887–905. doi: 10.1088/0031-9155/50/12/011
11. Zhong H, Kim J, Chetty IJ. Analysis of deformable image registration accuracy using computational modeling. *Med Phys* 2010; **37**: 970–9. doi: 10.1118/1.3302141
12. Fallone BG, Rivest DR, Riauka TA, Murtha AD. Assessment of a commercially available automatic deformable registration system. *J Appl Clin Med Phys* 2010; **11**: 3175.
13. Kashani R, Hub M, Kessler ML, Balter JM. Technical note: a physical phantom for assessment of accuracy of deformable alignment algorithms. *Med Phys* 2007; **34**: 2785–8. doi: 10.1118/1.2739812
14. Nithiananthan S, Brock KK, Daly MJ, Chan H, Irish JC, Siewerdsen JH. Demons deformable registration for CBCT-guided procedures in the head and neck: convergence and accuracy. *Med Phys* 2009; **36**: 4755–64. doi: 10.1118/1.3223631
15. Salguero FJ, Saleh-Sayah NK, Yan C, Siebers JV. Estimation of three-dimensional intrinsic dosimetric uncertainties resulting from using deformable image registration for dose mapping. *Med Phys* 2011; **38**: 343–53. doi: 10.1118/1.3528201
16. Bender ET, Hardcastle N, Tomé WA. On the dosimetric effect and reduction of inverse consistency and transitivity errors in deformable image registration for dose accumulation. *Med Phys* 2012; **39**: 272–80. doi: 10.1118/1.3666948
17. Wang H, Garden AS, Zhang L, Wei X, Ahamad A, Kuban DA, et al. Performance evaluation of an automatic anatomy segmentation algorithm on repeat or four-dimensional CT images using a deformable image registration method. *Int J Radiat Oncol Biol Phys* 2008; **72**: 210–19. doi: 10.1016/j.ijrobp.2008.05.008
18. Brock KK; Deformable Registration Accuracy Consortium. Results of a multi-institution deformable registration accuracy study (MIDRAS). *Int J Radiat Oncol Biol Phys* 2010; **76**: 583–96. doi: 10.1016/j.ijrobp.2009.06.031
19. La Macchia M, Fellin F, Amichetti M, Cianchetti M, Gianolini S, Paola V, et al. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. *Radiat Oncol* 2012; **7**: 160. doi: 10.1186/1748-717x-7-160
20. Mencarelli A, van Beek S, van Kranen S, Rasch C, van Herk M, Sonke JJ. Validation of deformable registration in head and neck cancer using analysis of variance. *Med Phys* 2012; **39**: 6879–84. doi: 10.1118/1.4760990
21. Thor M, Petersen JB, Bentzen L, Høyer M, Muren LP. Deformable image registration for contour propagation from CT to cone-beam CT scans in radiotherapy of prostate cancer. *Acta Oncol* 2011; **50**: 918–25. doi: 10.3109/0284186X.2011.577806
22. Liu H, Wu Q. Dosimetric and geometric evaluation of a hybrid strategy of offline adaptive planning and online image guidance for prostate cancer radiotherapy. *Phys Med Biol* 2011; **56**: 5045–62. doi: 10.1088/0031-9155/56/15/024
23. Varadhan R, Karangelis G, Krishnan K, Hui S. A framework for deformable image registration validation in radiotherapy clinical applications. *J Appl Clin Med Phys* 2013; **14**: 4066. doi: 10.1120/jacmp.v14i1.4066
24. Liu H, Wu Q. Evaluations of an adaptive planning technique incorporating dose feedback in image-guided radiotherapy of prostate cancer. *Med Phys* 2011; **38**: 6362–70. doi: 10.1118/1.3658567
25. Wen N, Kumarasiri A, Nurushev T, Burmeister J, Xing L, Liu D, et al. An assessment of PTV margin based on actual accumulated dose for prostate cancer radiotherapy. *Phys Med Biol* 2013; **58**: 7733–44. doi: 10.1088/0031-9155/58/21/7733
26. Hunter KU, Fernandes LL, Vineberg KA, McShan D, Antonuk AE, Cornwall C, et al. Parotid glands dose-effect relationships based on their actually delivered doses: implications for adaptive replanning in radiation therapy of head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2013; **87**: 676–82. doi: 10.1016/j.ijrobp.2013.07.040
27. Hub M, Thieke C, Kessler ML, Karger CP. A stochastic approach to estimate the uncertainty of dose mapping caused by uncertainties in b-spline registration. *Med Phys* 2012; **39**: 2186–92. doi: 10.1118/1.3697524
28. Yan C, Hugo G, Salguero FJ, Saleh-Sayah N, Weiss E, Sleeman WC, et al. A method to evaluate dose errors introduced by dose mapping processes for mass conserving deformations. *Med Phys* 2012; **39**: 2119–28. doi: 10.1118/1.3684951
29. Zhong H, Weiss E, Siebers JV. Assessment of dose reconstruction errors in image-guided radiation therapy. *Phys Med Biol* 2008; **53**: 719–36. doi: 10.1088/0031-9155/53/3/013
30. Murphy MJ, Salguero FJ, Siebers JV, Staub D, Vaman C. A method to estimate the effect of deformable image registration uncertainties on daily dose mapping. *Med Phys* 2012; **39**: 573–80. doi: 10.1118/1.3673772
31. Janssens G, de Xivry JO, Fekkes S, Dekker A, Macq B, Lambin P, et al. Evaluation of

nonrigid registration models for interfraction dose accumulation in radiotherapy. *Med Phys* 2009; **36**: 4268–76. doi: 10.1118/1.3194750
32. Huang TC, Liang JA, Dilling T, Wu TH, Zhang G. Four-dimensional dosimetry validation and study in lung radiotherapy using deformable image registration and Monte Carlo techniques. *Radiat Oncol* 2010; **5**: 45. doi: 10.1186/1748-717x-5-45
33. Niu CJ, Foltz WD, Velec M, Moseley JL, Al-Mayah A, Brock KK. A novel technique to enable experimental validation of deformable dose accumulation. *Med Phys* 2012; **39**: 765–76. doi: 10.1118/1.3676185
34. Yeo UJ, Taylor ML, Supple JR, Smith RL, Dunn L, Kron T, et al. Is it sensible to "deform" dose? 3D experimental validation of dose-warping. *Med Phys* 2012; **39**: 5065–72. doi: 10.1118/1.4736534
35. Juang T, Das S, Adamovics J, Benning R, Oldham M. On the need for comprehensive validation of deformable image registration, investigated with a novel 3-dimensional deformable dosimeter. *Int J Radiat Oncol Biol Phys* 2013; **87**: 414–21. doi: 10.1016/j.ijrobp.2013.05.045
36. Schultheiss T, Tomé WA, Orton CG. Point/counterpoint: it is not appropriate to "deform" dose along with deformable image registration in adaptive radiotherapy. *Med Phys* 2012; **39**: 6531–3. doi: 10.1118/1.4722968
37. Bookstein FL. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans Pattern Anal Mach Intell* 1989; **11**: 567–85. doi: 10.1109/34.24792
38. Wachowiak MP, Wang X, Fenster A, Peters TM. Compact support radial basis functions for soft tissue deformation. In: *IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004*. IEEE; 2004. pp. 1259–62.
39. Thirion JP. Image matching as a diffusion process: an analogy with Maxwell's demons. *Med Image Anal* 1998; **2**: 243–60. doi: 10.1016/S1361-8415(98)80022-4
40. Jena R, Kirkby NF, Burton KE, Hoole AC, Tan LT, Burnet NG. A novel algorithm for the morphometric assessment of radiotherapy treatment planning volumes. *Br J Radiol* 2010; **83**: 44–51. doi: 10.1259/bjr/27674581
41. Barker JL Jr, Garden AS, Ang K, O'Daniel JC, Wang H, Court LE, et al. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. *Int J Radiat Oncol Biol Phys* 2004; **59**: 960–70. doi: 10.1016/j.ijrobp.2003.12.024

APPENDIX A

Analysis for volume conservation of critical structures during the application of artificial deformations (forward and backward neck flexion) revealed an agreement of ≥99% for all structures. An estimation of the potential mean "worst-case scenario" dosimetric uncertainties owing to these volumetric changes, in our test cases, was calculated using the following equations:

$$\bar{e}_{\text{median}} = \frac{\sum_{i=1}^{N} \max(|D_{50\%} - D_{49.5\%}|, |D_{50\%} - D_{50.5\%}|)}{N}$$

$$\bar{e}_{\text{max}} = \frac{\sum_{i=1}^{N} |D_{\text{max}} - D_{1\%}|}{N}$$

$$\bar{e}_{\text{min}} = \frac{\sum_{i=1}^{N} |D_{\text{min}} - D_{99\%}|}{N}$$

where $\bar{e}_{\text{median}}$, $\bar{e}_{\text{max}}$ and $\bar{e}_{\text{min}}$ is the estimated "worst-case scenario" error in median, minimum and maximum dose to each organ, respectively. D_{min} is the minimum and D_{max} the maximum dose delivered to each structure. $D_{50\%}$, $D_{49.5\%}$, $D_{50.5\%}$, $D_{99\%}$, $D_{1\%}$ is the dose received by 50%, 49.5%, 50.5%, 99% and 1% of each organ's volume, respectively. N is the number of test cases, which in this study was 12.

Estimated potential dosimetric uncertainties in the test cases used, for the observed 1% difference in volume are summarized in Table A1. It is assumed that the potential uncertainty in the estimation of the mean dose to a certain organ owing to a 1% change in volume would be less than that of the median dose and was therefore not estimated.

Table A1. Estimated mean "worst-case scenario" errors in the calculation of median ($\bar{e}_{\text{median}}$), maximum ($\bar{e}_{\text{max}}$) and minimum ($\bar{e}_{\text{min}}$) dose to the organs and volumes under investigation

Region of interest	$\bar{e}_{\text{median}}$ (Gy)	$\bar{e}_{\text{max}}$ (Gy)	$\bar{e}_{\text{min}}$ (Gy)
Brain	0.002	0.013	–
Brainstem	0.005	0.011	–
Spinal cord	0.004	0.010	–
Contralateral parotid	0.001	0.008	–
Planning target volume	0.001	–	0.020