Published in final edited form as:

Int J Radiat Oncol Biol Phys. 2008 February 1; 70(2): 572–581.

Comparison of intensity-modulated radiotherapy planning based on manual and automatically generated contours using deformable image registration in four-dimensional computed tomography of lung cancer patients

ELISABETH WEISS, M.D.^{1,2,*}, KRISHNI WIJESOORIYA, Ph.D.³, VISWANATHAN RAMAKRISHNAN, Ph.D.⁴, and PAUL J. KEALL, Ph.D.⁵

¹ Department of Radiation Oncology, Virginia Commonwealth University, Richmond, VA, U.S.A

² Department of Radiation Oncology, University of Göttingen, Göttingen, Germany

³ Department of Radiation Oncology, University of Pittsburgh, Pittsburgh, PA, U.S.A

⁴ Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, U.S.A

⁵ Department of Radiation Oncology, Stanford University, Stanford, CA, U.S.A

Abstract

Purpose—To evaluate the implications of differences between manually drawn contours and contours automatically generated by deformable image registration (DIR) for 4D treatment planning.

Material and Methods—In 12 lung cancer patients intensity-modulated radiotherapy (IMRT) planning was performed for both manual contours and automatically generated (“auto”) contours in mid and peak expiration of four-dimensional computed tomography scans with the manual contours in peak inspiration serving as the reference for the displacement vector fields. Manual and auto plans were analysed with respect to their coverage of the manual contours which were assumed to represent the anatomically correct volumes.

Results—Auto contours were on average larger than manual contours by up to 9%. Objective scores, D_{2%} and D_{98%} of the PTV, homogeneity and conformity indices, and coverage of normal tissue structures (lungs, heart, esophagus, spinal cord) at defined dose levels were not significantly different between both plans (p=0.22–0.94). Differences were statistically insignificant for the generalized equivalent uniform dose of the PTV (p=0.19–0.94) and normal tissue complication probabilities for lung and esophagus (p=0.13–0.47). Dosimetric differences >2% or >1 Gy were more frequent in patients with auto/manual volume differences ≥10% (p=0.04).

Conclusion—The applied DIR algorithm produces clinically plausible auto contours in the majority of structures. At this stage clinical supervision of the auto contouring process is required and manual interventions may become necessary. Prior to routine use further investigations are required to particularly reduce imaging artifacts.

*corresponding author: Dr. E. Weiss, Department of Radiation Oncology, Virginia Commonwealth University, 401 College Street, PO Box 980058, Richmond, VA 23298, phone: 804-828-9463, fax: 804-828-6042, e-mail: eweiss@mcvh-vcu.edu.

Conflicts of interest notification:

None of the authors has any actual or potential conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Deformable image registration; automatic contouring; lung cancer; intensity modulated radiotherapy

Introduction

Imaging accounting for respiration-induced intrafractional motion, in particular four-dimensional computed tomography (4D CT), is becoming widely used for thoracic and abdominal radiotherapy¹. 4D CTs are valuable for determining tumor and normal tissue motion and deformation^{2, 3}, for safety margin definition in motion-encompassing treatment methods⁴, for planning respiratory-gated radiotherapy and 4D tumor tracking⁵.

To take advantage of the full information contained in 4D CT images for the planning of respiration-synchronized treatment techniques requires 3D structure delineation on the image sets of all respiratory phases. This increase in the contouring work load exceeds available time resources of qualified personnel. For each phase of a 4D CT scan approximately 1.5 to 3 hours or even longer are required, depending on slice thickness, number of contoured structures and tumor size. Automation of structure delineation is therefore required to make planning based on 4D CT information a workable process.

Several deformable image registration algorithms for use in CT images are available today, some of them have been applied to 4D CT data or to generate 4D CT images^{5–18}.

In the present study large diffeomorphic image registration was employed to automatically generate structures for all respiratory phases of 4D CTs. Large diffeomorphic image registration has been developed to automatically produce structures in the presence of non-rigid tissue motion^{19–21}. Deformable image registration assumes that each point in one CT corresponds to a point in a CT acquired at a different time or respiration phase.

In an analysis by Pevsner et al.²² this method was applied to predict tumor and lung deformation in the thorax in respiration-correlated CT data sets. The comparison of model-predicted and observer-drawn structures showed that the accuracy of automatically mapped structures was comparable to observer variability. The authors concluded that the automated image matching method had the potential to predict and quantify lung tumor motion in CT. In our own analysis²³ in addition to tumor volumes also manual and automatically generated normal tissue structures were analyzed volumetrically. Good agreement was observed, particularly for tumor volumes. While volumes and surfaces of defined manual and auto structures might not differ significantly and variations might be similar to the contouring variability of different observers, differences do exist. Pevsner et al.²² found that for GTVs surface discrepancies between observer-drawn and auto contours were larger than 5.1 mm in 10%.

A more comprehensive evaluation than the geometric analysis of individual structures is therefore necessary prior to clinical use. Treatment planning appears to be an appropriate method to test all relevant aspects of auto contours including volume, shape and position also relative to other structures in a clinical setting. Treatment planning based on auto contours and comparison to planning based on manual contours enables to determine clinical reliability and to establish confidence in the clinical use.

The aims of the current study were to analyze the dosimetric effect of auto contours for treatment planning in the thorax and to compare results to planning based on manual contours in 4D CTs.

Material and Methods

This study is based on 4D CTs of 12 consecutive patients with lung cancer. Tumor stages ranged from T1N0 to T3N2 and T4N0 with positive lymph nodes in 6 patients. Two patients had tumors located in the periphery of the lower lobes, one was centrally in the lower lobes, one was in the periphery in the upper lobes and eight tumors were in the central upper lobes of the lungs. The mean GTV volume was 49.0 cm³ (range 2.9 – 317.7 cm³). The average maximum center of mass motion during respiration for these tumors was 0.7 cm (range 0.1 – 2.4 cm).

IRB approval and patient consent were obtained for all involved patients

4D CT acquisition

All patients underwent respiration-correlated 4D CTs that were acquired on a 16-slice CT scanner in cine mode with a 2.5 mm slice thickness (GE Healthcare Technologies, Waukesha, WI). CT images were provided by the MD Anderson Cancer Center in Houston, TX. Respiratory motion was detected using a respiration-monitoring system (RPM, Varian Medical Systems, Palo Alto, CA). All images were retrospectively binned into 10 respiration-correlated phases T0 to T9 of equal length with T0 representing peak inspiration and T5 representing the position in the middle between two adjacent T0 phases which was usually in end expiration.

Manual Contouring

For manual contouring a commercially available planning system (Pinnacle, Version 6.2 and 7.7, Philips Medical Systems, Milpitas, CA) was used. Contouring was performed by four experienced medical personnel according to a detailed contouring protocol. The following structures were contoured for all phases and all patients: Gross tumor volume (GTV), lungs, heart, spinal cord and esophagus. GTV, heart and esophagus were segmented completely by hand. Lung contours were created using the Pinnacle software for auto contouring and edited manually to follow the visible lung borders. The spinal cord contour (not the spinal canal) was produced by placing a circle-shaped contour with a patient-dependent diameter to cover the spinal cord on each slice. This diameter remained the same for all slices and all phases in the image sets of the respective patient. The analysis of the manual contours has been described by Weiss et al.³

Auto Contouring

In this study the large deformation diffeomorphic image registration algorithm as developed at the University of North Carolina was used for deformable image registration and auto contour production. A more detailed description of this algorithm can be found in the references^{19–21}.

In order to map all points of one respiratory phase to the corresponding points in the other respiratory phases, in a first step the transformation accounting for variations of position and shape of structures during respiration was identified. The displacement vector fields from one respiratory phase to another were calculated in several iterations from coarser voxel scales to finer scales until convergence was reached. The displacement vector field for each patient in this study was based on the manual contours of the CT images in phase T0 (peak inspiration).

In a second step the transformation from the deformable image registrations was applied to the manual contours defined in the T0 phase. This resulted in the automatic generation of contours on subsequent respiratory phases of the 4D CT set.

A quantification of the geometric difference between manually defined structures and those automatically determined using the deformable image registration was performed for all 6 structures²³. The current work focuses on the dosimetric impact of the geometric differences.

Planning

Based on the manually contoured (“manual”) and automatically generated (“auto”) GTVs of the individual phases, manual and auto clinical target volumes (CTVs) and planning target volumes (PTVs) of the respective phases were created by adding an isotropic 8 mm margin to the GTVs in order to generate CTVs and by adding another 8 mm margin to the CTVs to produce PTVs. While for patient treatment, depending on tumor motion and the respiration management system, often anisotropic safety margins are chosen, for simplicity uniform margins were used in this plan comparison study.

Manual and auto spinal cord and esophagus contours were expanded by 5 mm each to obtain the respective manual and auto planning organ at risk volumes (PRVs). The total lung volume was defined as the sum of the right and left lung minus the GTV volume. A structure called skin-PTV containing the whole thorax volume except the PTV was created which was used to limit the high dose volumes outside the PTV during planning.

For treatment planning based on both the manual and auto contours two representative respiratory phases, phase 2 and 5, were selected:

- a. Phase 2 is an intermediate respiratory phase between in- and expiration and is potentially affected by hysteresis.
- b. Phase 5 has the largest motion trajectory in the majority of patients and therefore is assumed to be at the highest risk for discrepancies between manual and auto contours.

Intensity modulated radiotherapy (IMRT) planning was performed with the Pinnacle treatment planning system (version 7.9) using a collapsed cone convolution implementation of the superposition/convolution algorithm with inhomogeneity correction and DMPO (direct machine parameter optimization). In general, six 6-MV coplanar non-opposed photon beams were chosen with angles depending on the tumor location. The isocenters were defined as the centroids of the manual and auto GTVs for the respective plans. While beam angles remained the same for the plans based on both the manual and auto contour sets of one patient, beam weights and positions of the multi-leaf collimators (MLCs) varied as a consequence of plan optimization.

The total dose was 74 Gy in 37 fractions to the manual and auto PTVs. The objective of inverse planning was to deliver the prescribed dose to at least 95% of the PTV with a dose range not exceeding -10% and +20% of the prescribed dose^{24, 25}. Plan optimization used the same constraints for both the manual and the auto plans. The following constraints were applied in this study: The lung volume receiving 20 Gy or higher ($V_{20\text{Gy}}$) was not to exceed 30%, the heart volume receiving 40 Gy or higher ($V_{40\text{Gy}}$) should not exceed 50%. A dose of 55 Gy or higher ($V_{55\text{Gy}}$) should be limited to 30% of the esophagus PRV and the maximum dose ($D_{1\%}$) at the spinal cord PRV should not go beyond 45 Gy. Dose in the thorax outside the PTV ($D_{1\%}$) should be limited to 80 Gy.

Presently, manual contours are the standard in structure delineation. They are assumed to represent the actual anatomy. Therefore, while dose in plans created for auto contours (“auto plans”) was optimized using objectives based on the automatically generated contours, auto plans were evaluated for their coverage of the original manual contours. Auto plans were compared to the plans based on manual contours (“manual plans”) which served as a

benchmark. Figure 1 illustrates the contouring, planning and evaluation procedures performed in this study.

Analysis

The following criteria were analysed:

- Overall plan quality

This study is based on the assumption that both the manual and auto plans provide optimum plans for the respective manual or auto contour set with regard to planning objectives and dose constraints. To determine overall plan quality, the objective composite score CS was calculated and compared for both plans. The CS as a single metric is an appropriate tool for plan quality comparison that takes into account the actually achieved dose distribution for PTV and organs at risk after plan optimization, relative to the initially given constraints. The CS is the weighted sum of the plan subscores for individual anatomic structures. A plan subscore is a measure for the deviation from the specified dose, dose-volume, or dose-response objective for that structure. The greater the deviation, the greater the plan subscore for that structure. Monitor units (MUs) for both plans were compared as a factor contributing to cost benefit aspects.

- Comparison of manual and auto contour volumes

Volumes of all GTVs, PTVs and normal tissue structures (Lungs-GTV, Heart, Esophagus PRV and Cord PRV) were compared for both respiratory phases and both manual and auto contouring including an analysis of absolute volumes and ratios between manual and auto contours.

- Dose-volume histogram (DVH) analysis

For both the manual and the auto contour plans DVHs were analysed for the dose and volume coverage of the manual contours. Dose to the manual PTV was analysed comparing PTV D_{2%} (minimum dose in 2% of the PTV volume), D_{98%} and D_{50%}. Results were related to phase, tumor location and tumor size. In addition, a dose homogeneity index ($HI = D_{5\%}/D_{95\%}$) and a dose conformity index were calculated. The conformity index (CI) was defined as follows: $CI = CF \text{ (cover factor)} \times SF \text{ (spill factor)}$, where the CF was defined as the percentage of the PTV volume receiving at least 74 Gy and the SF as the volume of the PTV receiving at least 74 Gy relative to the total 74-Gy volume (see also Radiation Therapy and Oncology Group (RTOG) protocol 98-03). The closer the HIs and CIs were to 1, the better were dose homogeneity and conformity. DVHs for normal tissue were compared for different dose and volume parameters, among them the parameters used as constraints for plan optimization (Lungs-GTV V_{20Gy}, Lungs-GTV D_{50%}, Heart V_{40Gy}, Heart D_{50%}, Esophagus PRV V_{55Gy}, Cord PRV D_{1%}, Thorax-PTV D_{1%}). As an arbitrary measure for clinically relevant differences, volume variations >2% for specified parameters and differences of the spinal cord D_{1%} value >1Gy were investigated.

- Comparison of biological outcome parameters

Values for generalized equivalent uniform dose (gEUD) and normal tissue complication probabilities (NTCP) were calculated in Pinnacle using the DVHs for manual contours for both plans. Normal tissue complication probabilities (NTCPs) were calculated for the lung using the Lyman model²⁶ with a median toxic dose (TD₅₀) of 30.8 Gy, a steepness m of 0.37 and a volume component n of 0.99²⁷. To calculate the NTCP values for the esophagus, a TD₅₀ of 47 Gy, a steepness m of 0.36 and an n value of 0.69 were used²⁸.

- Cause and dosimetric effect of larger manual and auto volume differences

In order to assess the dosimetric effect of larger manual and auto volume differences, an analysis of individual structures with volume differences $\geq 10\%$ was performed with the intention to both identify the cause for the observed volume differences and to evaluate the dosimetric consequences for the structure itself and in case of the PTV for normal tissue.

A random effects ANOVA was used for the analysis. The independent variables in this analysis were method of contouring (manual versus auto contours), breathing position (mid inhalation or exhalation) and their interaction. To account for the subject to subject variability a random intercept statement was included. To investigate the significance of volume differences on dosimetric outcome, a Fisher's exact test was used. A p-value < 0.05 was considered significant.

Results

- Overall plan quality

The mean objective score CS was not significantly different between manual and auto contours for both phases (Table 1, $p=0.22-0.78$). CS values were slightly worse for phase T5 compared to T2 which can be attributed to smaller lung volumes during expiration. Average values of monitor units did not vary significantly between manual and auto plans ($p=0.28-0.62$).

- Comparison of manual and auto contour volumes

Figure 2 shows examples for manual and automatically generated contours for GTV and normal tissue structures in T2 and T5 in one patient. Volumes of GTVs, PTVs and normal tissue both for manual and auto contours are shown in tables 2a and 2b. In phase T2 and T5, auto contours were on average larger than manual contours by up to 9%. Differences in the magnitude of volumes were significant for lungs – GTV ($p<0.0001$).

- Dose-volume histogram (DVH) analysis

For the PTV on average no significant differences in all evaluated parameters were observed (Table 3a and b, $p=0.58-0.79$). In phase T2 and T5, $D_{50\%}$ differed by more than 1 Gy, but less than 2.5 Gy in two patients in phase T2 (small peripheral and small central tumor) and one patient in T5 (large peripheral tumor). Variations in the minimum dose values to the PTV $D_{98\%}$ were as expected larger than 2 Gy in 5 patients each in T2 and T5. In two patients differences in $D_{98\%}$ up to 25 Gy were observed. These differences were not related to any location- or size-specific tumor characteristics.

Differences were not significantly different for all analyzed normal tissue parameters (table 4a and b, $p=0.41-0.94$). Average values were slightly worse for auto compared to manual plans with the exception of esophagus PRV $V_{55\text{Gy}}$ in T2 and T5, cord PRV $D_{1\%}$ in T2 and lungs-GTV $D_{50\%}$ in T5.

Dosimetric differences $> 2\%$ or > 1 Gy were found for altogether 13 structures: In two patients for lung $V_{20\text{Gy}}$ (one in T2 and one in T5), two patients for esophagus PRV $V_{55\text{Gy}}$ (one in T2 and one in T5), two patients for heart $V_{40\text{Gy}}$ in T2, seven patients for cord PRV $D_{1\%}$ (four patients in T2 and three patients in T5). In addition, dose differences larger than 1 Gy were also observed for the $D_{1\%}$ of thorax-PTV in phase T2 and T5 for 6 patients each. A plan comparison for a large central tumor is shown in figure 3.

- Comparison of biological outcome parameters

Generalized EUD values were higher for T2 and T5 for manual plans compared to auto plans ($p=0.19-0.94$) (table 5a and b). The estimated complication risk was slightly lower with manual plans for lung in T2 and T5 and for esophagus in T2 ($p=0.13-0.47$).

- Cause and dosimetric effect of larger manual and auto volume differences

In individual patients volume differences $\geq 10\%$ were observed for the ratio of auto/manual volumes. The appendix shows the individual data of these patients including the dosimetric effect of these volume differences. Four patients had volume difference $\geq 10\%$ in the PTVs. These were two patients with small peripheral tumors and two patients with central tumors. Differences in PTV volumes $\geq 10\%$ resulted in dose differences of up to 3 Gy at the PTV $D_{98\%}$ level. The impact of PTV volume differences on normal tissue dose is difficult to judge, since in two patients with PTV volume differences $\geq 10\%$ also larger normal tissue volume differences were observed which per se could result in dosimetric changes. In patient 12, however, PTV volume differences potentially resulted in an increase in Lungs-GTV V_{20Gy} of 3%. Differences in the lungs-GTV volumes of this patient were 2.6%.

Dose variations in normal tissue structures with volume differences $\geq 10\%$ were below 2% and 1 Gy at specified tolerance dose levels, with four exceptions, see appendix. The odds for having dosimetric differences larger than 1 Gy or more than 2% when having volume differences $\geq 10\%$ is approximately 4 times higher than when having volume differences $< 10\%$ ($p=0.04$).

Discussion

In the present study no statistically significant differences were observed for all investigated dosimetric parameters both in phase T2 and T5. This indicates that the applied deformable image registration algorithm produces in general reliable auto contours both in the peak expiration phase which is potentially most affected by errors in the displacement vector field and an intermediate phase where hysteresis might influence the accuracy of auto contours.

The analysis of individual patients showed differences in auto/manual volumes $\geq 10\%$ in the majority of patients both for PTV and normal tissue. Volume differences $\geq 10\%$ were observed both in patients with central and peripheral, and upper as well as lower lobe tumors. In none of these patients did the use of auto contours lead to changes in normal tissue doses that exceeded the given tolerance levels. The effect of volume differences on dose was mostly less than 2% or lower than 1 Gy and was therefore assumed tolerable from the clinical view. PTV volume differences $\geq 10\%$ were noticed in two patients with small peripheral tumors of the lower lobes for both T2 and T5 phases in addition to two patients with central tumors in T5 phase. The observed volume differences between manual and auto contours in these 4 patients translated into $D_{98\%}$ differences of > 2 Gy in two patients in phase T2. The $D_{98\%}$ value represents the minimum dose in the PTV volume and is therefore the parameter that is affected most by a sharp dose gradient in IMRT planning and by changes in the PTV location relative to the external coordinate system. The observed differences in $D_{98\%}$ of the PTV as well as potential increases of normal tissue dose beyond the tolerance require clinical supervision of the auto contouring process. Depending on individual findings manual intervention might be required.

Small peripheral tumors are particularly prone to imaging artifacts in 4D CT scans, since respiration irregularities in areas with high respiration-related motion such as the inferior and peripheral parts of the lungs lead to binning errors and inaccurate tumor visualization, resulting in over- or underestimation of the actual anatomical volume³. In this study the maximum respiration-related tumor motion in the 9 patients with volume differences $\geq 10\%$ was higher than the average tumor motion of the 3 patients without large volume differences (8.0 cm versus 5.7 cm, n.s.). While for manual contours there are no established ways to correct for imaging artifacts during contouring, using deformable image registration a reduction of binning errors and a smoothing effect in auto contours were observed which level out extremes in volume variations that are physiologically not plausible. Deformable image registration therefore has

the potential to lead to anatomically more consistent volumes in all breathing phases in the presence of artifacts induced by irregular motion^{12, 23}.

While auto contours were clinically plausible in the vast majority of cases in this study, differences to manual contours were observed in the presence of manual contouring errors, i.e. when manual contours did not encompass the same anatomical structures in the CT sets of all respiration phases²³. Contouring errors occurred mostly during heart contouring where the inferior vena cava was included inconsistently in the heart contour of the different respiration phases.

Most differences between auto and manual contours were found in the superior-inferior direction that is also affected most by respiration-induced changes of organ length. Particularly for the inferior end of the esophagus at the gastroesophageal junction, auto contours often interpolated between the manual reference structure in T0 and the manual structure of the respective phase T2 or T5 resulting in structures that were longer for the auto than the manual contour.

The ratio of auto and manual contour volumes might vary with the choice of the reference phase for calculation of the displacement vector field (DVF). In inspiration the largest structures in the thorax, i.e. the lungs, have their maximum volume, while other structures do not change at all with respiration. If inspiration is chosen as the reference phase, DVFs have to account for different amounts of volume changes during expiration and might in an averaging effect either overestimate volumes that are unaffected by respiration or underestimate the amount of volume changes in the presence of particularly large changes like the inferior part of the esophagus. In patient 1 and 6 (appendix) motion of the inferior border of the lungs differed from inferior esophagus motion by 10 mm. In addition, phase shifts for the maximum organ motion of lungs and esophagus were observed which might influence DVFs and result in discrepancies between auto and manual contours. Another reason for the observed variations between manual and auto contours might be that deformable registration image similarity is balanced with constraints on deformation. If the weighting of the image similarity were higher, perhaps there would have been better agreement for the esophagus, however, this would also affect the similarity of the other structures in the study.

The automatic generation of structures for the radiotherapy planning process, particularly for respiration-correlated imaging has increasingly been investigated. Different from the use of deformable image registration in adaptive radiotherapy settings where imaging is repeated to modify radiotherapy according to fraction-specific variations, deformable image registration in 4D CTs is less subject to variations. Technical imaging parameters as well as patient positioning are identical during the 4D image acquisition process with the major uncertainty being respiration motion and intrafraction changes in positioning.

For image registration of 4D CTs registration accuracies between 2 and 3 mm have been described using B-spline registration^{10, 12}. In a comparison of manual and auto contours in lung tumor patients using the same algorithm as the one used in the present study, Pevsner et al.²² measured a mean difference of 2.1 mm between the GTV surfaces created by DIR and by manual contouring. In our own study a mean difference in manual and auto GTV surfaces of approximately 1.3 mm was observed for expiration²³.

While manual contouring has long been accepted as the gold standard to define organ and tumor volumes, several interobserver delineation studies have shown wide variations particularly in lung cancer which might exceed potential inaccuracies of automatically generated contours^{29–32}. Pevsner et al.²² found comparable results for the mean difference between auto and manual contours (2.9 mm) versus interobserver discrepancies in manual contour delineation (2.8 mm).

Considering the documented accuracy of DIR algorithms and the dosimetric results from the present study, the applied DIR algorithm appears to be useful for supervised clinical applications in 4D contouring and radiotherapy planning using respiration management³³. Currently, one minimization with multiscale iterations at different coarseness levels requires approximately 12 to 15 minutes of computer time per phase. This results in a huge reduction in contouring time saving hours of physicians' and dosimetrists' time resources. The registration process could in principle be fully automated, with computations beginning as soon as the 4D CT scan is obtained and be ready for use two hours after the scan, during which time the manual contouring on the reference phase could be performed.

Prior to routine use a reduction of imaging artifacts appears necessary to improve the reliability of manual structure delineation and to reduce propagation of systematic auto contouring errors which are induced by suboptimal manual contours in the reference phase for DIR.

Acknowledgements

The authors acknowledge Drs. Lei Dong and Radhe Mohan from the UT MD Anderson Cancer Center for supplying the CT data used for this study and Dr. Sarang Joshi for providing the deformable image registration algorithm. We thank Dill Vaughn, Virginia Robertson and Drew Dill for substantially contributing to the contouring work on the 4D CT data sets and Dr. Mirek Fatyga for support with scripting for plan evaluation. This work was partially supported by NCI R01 CA 93626.

References

1. Keall PJ, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Med Phys* 2006;33:3874–3900. [PubMed: 17089851]
2. Mageras GS, Pevsner A, Yorke ED, et al. Measurement of lung tumor motion using respiration-correlated CT. *Int J Radiat Oncol Biol Phys* 2004;60:933–941. [PubMed: 15465212]
3. Weiss E, Wijesooriya K, Dill SV, et al. Tumor and normal tissue motion in the thorax during respiration: Analysis of volumetric and positional variations using 4D CT. *Int J Radiat Oncol Biol Phys* 2007;67:296–307. [PubMed: 17189078]
4. Wolthaus JW, Schneider C, Sonke JJ, et al. Mid-ventilation CT scan construction from four-dimensional respiration-correlated CT scans for radiotherapy planning of lung cancer patients. *Int J Radiat Oncol Biol Phys* 2006;65:1560–1571. [PubMed: 16863933]
5. Keall PJ, Joshi S, Vedam SS, et al. Four-dimensional radiotherapy planning for DM-LC-based respiratory motion tracking. *Med Phys* 2005;32:942–951. [PubMed: 15895577]
6. Brock KK, Dawson LA, Sharpe MB, et al. Feasibility of a novel deformable image registration technique to facilitate classification, targeting, and monitoring of tumor and normal tissue. *Int J Radiat Oncol Biol Phys* 2006;64:1245–1254. [PubMed: 16442239]
7. Huang TC, Zhang G, Guerrero T, et al. Semi-automated CT segmentation using optic flow and Fourier interpolation techniques. *Comput Methods Programs Biomed* 2006;84:124–134. [PubMed: 17027116]
8. Lu W, Chen ML, Olivera GH, et al. Fast free-form deformable registration via calculus of variations. *Phys Med Biol* 2004;49:3067–3087. [PubMed: 15357182]
9. Lu W, Olivera GH, Chen Q, et al. Automatic re-contouring in 4D radiotherapy. *Phys Med Biol* 2006;51:1077–1099. [PubMed: 16481679]
10. Rietzel E, Chen GT. Deformable registration of 4D computed tomography data. *Med Phys* 2006;33:4423–4430. [PubMed: 17153421]
11. Sarrut D, Boldea V, Miguet S, et al. Simulation of four-dimensional CT images from deformable registration between inhale and exhale breath-hold CT scans. *Med Phys* 2006;33:605–617. [PubMed: 16878564]
12. Schreiber E, Chen GT, Xing L. Image interpolation in 4D CT using a BSpline deformable registration model. *Int J Radiat Oncol Biol Phys* 2006;64:1537–1550. [PubMed: 16503382]
13. Wang H, Dong L, O'Daniel J, et al. Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. *Phys Med Biol* 2005;50:2887–2905. [PubMed: 15930609]

14. Coselmon MM, Balter JM, McShan DL, et al. Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines. *Med Phys* 2004;31:2942–2948. [PubMed: 15587645]
15. Ragan D, Starkschall G, McNutt T, et al. Semiautomated four-dimensional computed tomography segmentation using deformable models. *Med Phys* 2005;32:2254–2261. [PubMed: 16121580]
16. Guerrero T, Sanders K, Noyola-Martinez J, et al. Quantification of regional ventilation from treatment planning CT. *Int J Radiat Oncol Biol Phys* 2005;62:630–634. [PubMed: 15936537]
17. Guerrero T, Zhang G, Huang TC, et al. Intrathoracic tumour motion estimation from CT imaging using the 3D optical flow method. *Phys Med Biol* 2004;49:4147–4161. [PubMed: 15470929]
18. Zhang T, Orton NP, Tome WA. On the automated definition of mobile target volumes from 4D-CT images for stereotactic body radiotherapy. *Med Phys* 2005;32:3493–3502. [PubMed: 16370433]
19. Christensen GE, Rabbitt RD, Miller MI. 3D brain mapping using a deformable neuroanatomy. *Phys Med Biol* 1994;39:609–618. [PubMed: 15551602]
20. Foskey M, Davis B, Goyal L, et al. Large deformation three-dimensional image registration in image-guided radiation therapy. *Phys Med Biol* 2005;50:5869–5892. [PubMed: 16333161]
21. Joshi S, Pizer S, Fletcher PT, et al. Multiscale deformable model segmentation and statistical shape analysis using medial descriptions. *IEEE Trans Med Imaging* 2002;21:538–550. [PubMed: 12071624]
22. Pevsner A, Davis B, Joshi S, et al. Evaluation of an automated deformable image matching method for quantifying lung motion in respiration-correlated CT images. *Med Phys* 2006;33:369–376. [PubMed: 16532942]
23. Wijesooriya K, Weiss E, Dill V, et al. Quantifying the properties and accuracy of a deformable image registration algorithm for 4D treatment planning. *Med Phys* 2006;33:2186.
24. Liu HH, Wang X, Dong L, et al. Feasibility of sparing lung and other thoracic structures with intensity-modulated radiotherapy for non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys* 2004;58:1268–1279. [PubMed: 15001272]
25. Murshed H, Liu HH, Liao Z, et al. Dose and volume reduction for normal lung using intensity-modulated radiotherapy for advanced-stage non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys* 2004;58:1258–1267. [PubMed: 15001271]
26. Lyman JT. Complication probability as assessed from dose-volume histograms. *Radiat Res Suppl* 1985;8:S13–19. [PubMed: 3867079]
27. Seppenwoolde Y, Lebesque JV, de Jaeger K, et al. Comparing different NTCP models that predict the incidence of radiation pneumonitis. Normal tissue complication probability. *Int J Radiat Oncol Biol Phys* 2003;55:724–735. [PubMed: 12573760]
28. Belderbos J, Heemsbergen W, Hoogeman M, et al. Acute esophageal toxicity in non-small cell lung cancer patients after high dose conformal radiotherapy. *Radiother Oncol* 2005;75:157–164. [PubMed: 15890421]
29. Giraud P, Elles S, Helfre S, et al. Conformal radiotherapy for lung cancer: different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists. *Radiother Oncol* 2002;62:27–36. [PubMed: 11830310]
30. Senan S, van Sornsen de Koste J, Samson M, et al. Evaluation of a target contouring protocol for 3D conformal radiotherapy in non-small cell lung cancer. *Radiother Oncol* 1999;53:247–255. [PubMed: 10660205]
31. Steenbakkers RJ, Duppen JC, Fitton I, et al. Reduction of observer variation using matched CT-PET for lung cancer delineation: a three-dimensional analysis. *Int J Radiat Oncol Biol Phys* 2006;64:435–448. [PubMed: 16198064]
32. Van de Steene J, Linthout N, de Mey J, et al. Definition of gross tumor volume in lung cancer: inter-observer variability. *Radiother Oncol* 2002;62:37–49. [PubMed: 11830311]
33. Sarrut D. Deformable registration for image-guided radiation therapy. *Z Med Phys* 2006;16:285–297. [PubMed: 17216754]

Figure 1.

Flowchart of contouring, planning and analysis procedures performed in this study.

Figure 2.

Examples for manual and auto contours for phases T2 and T5 in a patient with a large lung tumor in the lower parts of the right lung. Manual contours are represented by lines, corresponding auto contours are in colorwash. Contours for GTV (red line, purple colorwash), lungs (orange), heart (blue), esophagus (green) and cord (yellow) are shown.

Figure 3.

Plan comparison for a large central lung tumor. Dose distribution of the manual plan (a) and the auto plan (b). The manual planning target volume (shaded red), the 74 Gy (blue line), 40 Gy (green line) and 20 Gy (pink line) isodose curves are shown. Figure 3c shows the dose volume histogram for the manually drawn contours with dashed lines representing the manual plan and solid lines the auto plan.

Table 1

Objective scores and monitor units for both types of plans and both respiratory phases.

	Plan for manual contour	Plan for auto contour	Difference auto - manual plan	
	Mean	Mean	Mean	SD
CS T2	0.096	0.079	-0.017	0.04
CS T5	0.109	0.114	0.0046	0.05
MU T2	228.70	222.05	-6.65	45.62
MU T5	240.04	251.56	11.52	34.30

CS: objective score; MU: monitor unit; SD: standard deviation

Appendix

Dosimetric effect for differences between auto and manual contours $\geq 10\%$, including potential causes for volume differences from imaging and deformable image registration

Patient # and location of tumor	Structure and phase	% of auto volume by which manual volumes are smaller	Dose difference for specified parameters at prescription or tolerance level	Causes for volume differences from imaging and deformable image registration
1 Small peripheral tumor in lower lung	PTV T2	12	D _{98%} 3.01 Gy	Imaging artifact with double contour of GTV. Auto contour covers both artifact and correct tumor, error in caudal direction.
	PTV T5	23	D _{98%} 0.75 Gy	Auto PTV larger in caudal direction as a result of incorrect DVF based on imaging artifacts in other phases.
	Lungs- GTV T2	11	V _{20Gy} 8.69 %	Imaging artifact with double contour of diaphragm in 7 of 9 phases, right auto lung extends further caudal than manual contour similar to reference, extends about 1 cm into liver.
	Esophagus PRV T2	18	V _{55Gy} 0 %	Auto esophagus follows reference and therefore extends further caudal than manual esophagus T2 which is shorter due to expiration.
	Esophagus PRV T5	11	V _{55Gy} 0 %	See comment for T2.
	Heart T5	13	V _{40Gy} 0.34 %	Auto heart larger in caudal direction, interpolates with reference contour, manual contouring error in T5 resulting in reduced volume compared to reference contour.
2 Left central tumor	Esophagus PRV T2	13	V _{55Gy} 2.56 %	Auto esophagus interpolates between reference and T2 contour on transverse plane. Reference larger due to more endoesophageal air.
3 Right central tumor in upper lung	PTV T5	11	D _{98%} 0.49 Gy	In transverse plane good agreement, auto PTV more caudal interpolating between reference and manual PTV in T5.
	Heart T5	11	V _{40Gy} 0.23 %	In transverse plane good agreement, auto heart more caudal interpolating between reference and manual heart in T5.
	Cord PRV T2	13	D _{1%} 0.20 Gy	In several slices larger than reference and manual T2 contour.
	Cord PRV T5	13	D _{1%} 1.33 Gy	In several slices larger than reference and manual T5 contour.
4 Left central tumor in upper lung	Lungs- GTV T5	10	V _{20Gy} 0 %	Both auto lungs extend further caudal than manual contours in same phase and are similar to reference, therefore extend between 1–2 cm into diaphragm.
6 Left central tumor in upper lung	Esophagus PRV T2	14	V _{55Gy} 0.42 %	Good agreement in transverse plane, although auto contour follows reference shape more than T2 esophagus contour; in caudal direction auto contour longer by 4 slices.
	Esophagus PRV T5	10	V _{55Gy} 0.66 %	See comment for T2 esophagus; auto contour longer by 3 slices in caudal direction.
	Heart T5	11	V _{40Gy} 0 %	In transverse plane good agreement, auto heart longer by 1 caudal slice interpolating between reference and manual heart in T5, manual contouring error resulting in reduced volume compared to reference contour.
	Cord PRV T2	30	D _{1%} 0.93 Gy	Contouring error in T0, auto contour follows the larger reference structure rather than manual cord T2.
	Cord PRV T5	29	D _{1%} 1.97 Gy	See comment for T2.
7 Right peripheral tumor in lower lung	Heart T5	17	V _{40Gy} 0 %	Manual heart larger in T0 than T5. Auto heart close to reference in cranial and caudal direction, transverse good agreement.
9 Right central tumor	PTV T5	12	D _{98%} 0.24 Gy	Auto contour interpolates between reference and T5 PTV.
10 Left central tumor	Heart T5	17	V _{40Gy} 1.72 %	Auto contour longer than manual in craniocaudal direction, good agreement in transverse plan.
12 Right peripheral tumor in lower lobe	PTV T2	21	D _{98%} 2.28 Gy	Auto contour is longer than manual contour in caudal direction by 2 slices interpolating between reference and manual T2 PTV also in transverse planes.
	PTV T5	17	D _{98%} 0.82 Gy	Auto contour is longer than manual contour in caudal direction by 4 slices interpolating between reference and manual T5 PTV.

V: volume receiving the specified dose or higher; D: minimum dose in respective percentage of volume; GTV: gross tumor volume; PTV: planning target volume; PRV: planning organ at risk volume; reference: respective manual structure in T0.

Table 2

Table 2a. Volumes of manual and auto contours in phase T2.				
Volumes	Manual contour T2	Auto contours T2	Difference auto contour T2 -manual contour T2	
	Mean [cm³]	Mean [cm³]	Mean [cm³]	SD [cm³]
GTV	47.44	49.77	2.14	4.04
PTV	287.57	291.67	3.78	13.81
Lungs-GTV	2691.83	2687.29	72.73	73.79
Heart	633.18	650.50	15.99	24.68
Esophagus PRV	98.05	100.14	1.93	5.63
Cord PRV	76.65	83.03	5.88	3.85
Table 2b. Volumes of manual and auto contours in phase T5.				
Volumes	Manual contour T5	Auto contours T5	Difference auto contour T5 -manual contour T5	
	Mean [cm³]	Mean [cm³]	Mean [cm³]	SD [cm³]
GTV	51.73	54.74	3.00	5.08
PTV	292.42	296.31	3.60	15.13
Lungs-GTV	2427.52	2554.27	116.99	73.44
Heart	623.52	667.39	40.49	40.75
Esophagus PRV	99.83	100.92	1.00	5.07
Cord PRV	76.52	82.98	5.96	3.82

GTV: gross tumor volume; PTV: planning target volume; PRV: planning organ at risk volume; SD: standard deviation

Table 3

Table 3a. Dose volume histogram (DVH) analysis for the planning target volume (PTV) in phase T2				
PTV T2	Plan for manual contour	Plan for auto contour	Difference auto - manual plan	
	Mean	Mean	Mean	SD
D _{50%} [Gy]	74.67	74.70	0.03	0.91
D _{2%} [Gy]	78.28	78.23	-0.05	2.52
D _{98%} [Gy]	66.94	65.53	-1.29	2.91
HI	1.116	1.111	-0.005	0.054
CI	0.554	0.562	0.007	0.10
Table 3b. Dose volume histogram (DVH) analysis for the planning target volume (PTV) in phase T5.				
PTV T5	Plan for manual contour	Plan for auto contour	Difference auto - manual plan	
	Mean	Mean	Mean	SD
D _{50%} [Gy]	74.60	74.48	-0.11	0.53
D _{2%} [Gy]	78.29	78.61	0.29	2.32
D _{98%} [Gy]	64.97	64.45	-0.48	10.45
HI	1.126	1.133	0.006	0.085
CI	0.550	0.504	-0.047	0.157

D: minimum dose in respective percentage of volume; HI: homogeneity index; CI: conformity index; SD: standard deviation

Table 4

Table 4a. Dose volume histogram (DVH) analysis for normal tissue in phase T2.				
Structure T2	Plan for manual contour	Plan for auto contour	Difference auto - manual plan	
	Mean	Mean	Mean	SD
Lungs-GTV V _{20Gy} [%]	24.12	24.87	0.69	2.61
Lungs-GTV D _{50%} [Gy]	4.10	4.48	0.35	1.02
Heart V _{40Gy} [%]	10.58	10.62	0.04	2.16
Heart D _{50%} [Gy]	7.45	8.09	0.58	2.37
Esophagus PRV V _{55Gy} [%]	13.62	13.17	-0.42	1.22
Cord PRV D _{1%} [Gy]	38.25	38.06	-0.17	1.04
Thorax-PTV D _{1%} [Gy]	64.95	66.13	1.08	2.10
Table 4b. Dose volume histogram (DVH) analysis for normal tissue in phase T5.				
Structure T5	Plan for manual contour	Plan for auto contour	Difference auto - manual plan	
	Mean	Mean	Mean	SD
Lungs-GTV V _{20Gy} [%]	23.92	24.04	0.11	1.17
Lungs-GTV D _{50%} [Gy]	4.99	4.79	-0.18	0.99
Heart V _{40Gy} [%]	10.33	11.08	0.68	1.83
Heart D _{50%} [Gy]	7.82	7.83	0.01	1.11
Esophagus PRV V _{55Gy} [%]	14.75	14.50	-0.23	1.16
Cord PRV D _{1%} [Gy]	36.38	36.45	0.06	1.05
Thorax-PTV D _{1%} [Gy]	65.67	65.85	0.17	1.57

V: volume receiving the specified dose or higher; D: minimum dose in respective percentage of volume; SD: standard deviation; GTV: gross tumor volume; PTV: planning target volume; PRV: planning organ at risk volume

Table 5

Table 5a. Biological outcome parameters for phase 2.				
T2	Plan for manual contour	Plan for auto contour	Difference auto – manual plan	
	Mean	Mean	Mean	SD
gEUD [Gv]	74.39	74.37	−0.02	0.87
NTCP lung [%]	7.91	8.5	0.54	0.19
NTCP esophagus [%]	12.67	13.08	0.38	1.85
Table 5b. Biological outcome parameters for phase 5.				
T5	Plan for manual contour	Plan for auto contour	Difference auto – manual plan	
	Mean	Mean	Mean	SD
gEUD [Gv]	74.37	74.01	−0.33	0.86
NTCP lung [%]	9.25	9.83	0.53	1.61
NTCP esophagus [%]	14.08	13.33	−0.69	1.93

SD: standard deviation; gEUD: generalized equivalent uniform dose; NTCP: normal tissue complication probability