Effect of deformable registration on the dose calculated in radiation therapy planning CT scans of lung cancer patients[a)]

Alexandra R. Cunliffe
Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637

Clay Contee
Department of Radiation and Cellular Oncology, The University of Chicago,
5841 South Maryland Avenue, Chicago, Illinois 60637

Samuel G. Armato III, Bradley White, and Julia Justusson
Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637

Renuka Malik and Hania A. Al-Hallaq[b)]
Department of Radiation and Cellular Oncology, The University of Chicago,
5841 South Maryland Avenue, Chicago, Illinois 60637

Purpose: To characterize the effects of deformable image registration of serial computed tomography (CT) scans on the radiation dose calculated from a treatment planning scan.
Methods: Eighteen patients who received curative doses ($\geq$60 Gy, 2 Gy/fraction) of photon radiation therapy for lung cancer treatment were retrospectively identified. For each patient, a diagnostic-quality pretherapy (4–75 days) CT scan and a treatment planning scan with an associated dose map were collected. To establish correspondence between scan pairs, a researcher manually identified anatomically corresponding landmark point pairs between the two scans. Pretherapy scans then were coregistered with planning scans (and associated dose maps) using the demons deformable registration algorithm and two variants of the Fraunhofer MEVIS algorithm ("Fast" and "EMPIRE10"). Landmark points in each pretherapy scan were automatically mapped to the planning scan using the displacement vector field output from each of the three algorithms. The Euclidean distance between manually and automatically mapped landmark points (d_E) and the absolute difference in planned dose ($|\Delta D|$) were calculated. Using regression modeling, $|\Delta D|$ was modeled as a function of d_E, dose (D), dose standard deviation (SD_{dose}) in an eight-pixel neighborhood, and the registration algorithm used.
Results: Over 1400 landmark point pairs were identified, with 58–93 (median: 84) points identified per patient. Average $|\Delta D|$ across patients was 3.5 Gy (range: 0.9–10.6 Gy). Registration accuracy was highest using the Fraunhofer MEVIS EMPIRE10 algorithm, with an average d_E across patients of 5.2 mm (compared with >7 mm for the other two algorithms). Consequently, average $|\Delta D|$ was also lowest using the Fraunhofer MEVIS EMPIRE10 algorithm. $|\Delta D|$ increased significantly as a function of d_E (0.42 Gy/mm), D (0.05 Gy/Gy), SD_{dose} (1.4 Gy/Gy), and the algorithm used ($\leq$1 Gy).
Conclusions: An average error of <4 Gy in radiation dose was introduced when points were mapped between CT scan pairs using deformable registration, with the majority of points yielding dose-mapping error <2 Gy (approximately 3% of the total prescribed dose). Registration accuracy was highest using the Fraunhofer MEVIS EMPIRE10 algorithm, resulting in the smallest errors in mapped dose. Dose differences following registration increased significantly with increasing spatial registration errors, dose, and dose gradient (i.e., SD_{dose}). This model provides a measurement of the uncertainty in the radiation dose when points are mapped between serial CT scans through deformable registration. © *2015 American Association of Physicists in Medicine.*
[http://dx.doi.org/10.1118/1.4903267]

1. INTRODUCTION

For patients who receive radiation therapy (RT) treatment for lung cancer, radiation-induced toxicities including radiation pneumonitis (RP) and, later, radiation fibrosis may develop when radiation-sensitive normal lung tissue is damaged. Approximately 50%–100% of patients who receive lung radiation therapy will experience radiologically evident normal tissue inflammation, with up to 20% diagnosed with symptomatic RP.[1,2] RP incidence is highly dependent on the delivered dose and the volume of normal lung tissue irradiated.[3] Thus, many studies have investigated metrics for RP prediction that are based on dose-volume histograms generated from a patient's treatment plan.[4,5] Discrepancies in these findings, however,

may be due to variability in treatment plan design or variable patient sensitivity to radiation.[5] As a result, no factor has been identified as consistently predictive of RP development.

Several studies have adopted a more individualized approach to assess radiation-induced lung tissue damage based on a patient's post-RT scan. Zhang *et al.*[6] demonstrated the feasibility of registering radiation dose with standardized uptake value (SUV) maps of fluorine-18-fluorodeoxyglucose positron emission tomography (^{18}F-FDG PET)/computed tomography (CT) scans acquired after RT to assess normal tissue damage. Phernambucq *et al.*[7] measured a significant increase in regional CT scan density with increasing radiation dose for scans acquired >3 months following lung RT in 25 patients. To obtain voxel-by-voxel association between planned radiation dose and post-RT scans, both groups used deformable image registration to align radiation dose maps associated with the treatment planning scan with post-RT CT or PET/CT scans.

Deformable image registration has been widely studied for use in lung scan coregistration to accommodate the nonuniform changes introduced by patient respiration. Several multi-institutional studies have demonstrated high accuracy (i.e., <1 mm average error) in lung CT scan coregistration.[8,9] Our past experience with the Plastimatch demons deformable registration algorithm[10] showed that average registration errors of less than 0.5 mm were present between serial diagnostic CT scans obtained from healthy patients, compared with 4.8-mm average errors using rigid registration.[11] Even the best algorithms, however, introduce errors during image registration; a multi-institutional study by Brock[9] found that the nine most accurate registration algorithms in the study introduced errors as high as 5 mm when coregistering four-dimensional CT (4D-CT) lung scans. Registration may be further complicated when large differences between CT scans exist due to differences in patient positioning, breathing phase, or disease status, thus resulting in decreased registration accuracy. When deformable registration is used to map dose between scans, registration errors yield errors in the dose associated with the mapped location. Several studies have evaluated the dose-mapping error that results from inaccurate deformation of lung CT scans.[12,13] The purpose of these studies was to incorporate respiratory changes during treatment planning; thus, these studies involved a single 4D-CT case as a demonstration of feasibility.

Treatment planning CT scans and dose maps could alternatively be registered with diagnostic CT scans acquired outside of the RT treatment dates, which would allow for correlation between planned dose and radiation-induced changes in lung tissue. The current study evaluated the dose-mapping error when the planning scans of 18 lung cancer patients were coregistered with diagnostic-quality CT scans acquired prior to RT. Scans were acquired during different imaging sessions that used different scanners, imaging parameters, patient setup, and respiratory phase; each of these factors can complicate the registration process. This study characterized how dose-mapping error was affected by local dose, dose gradient, the extent of registration accuracy, and the registration algorithm used. The goal was to determine the individual contributions of multiple factors to the error introduced when deformable registration was used to map planned radiation dose between serial CT scans.

2. METHODS

2.A. Patient database

Twenty patients who underwent curative-intent RT (≥ 60 Gy) for lung cancer at The University of Chicago Medicine between 2007 and 2011 were identified under IRB approval. These patients were randomly selected from a previously compiled database[14] of 25 patients who had pre- and post-RT diagnostic CT scans and treatment planning scans available at our institution. Patients were treated with 6 MV or a mixture of 6 and 18 MV photons while immobilized using custom alpha cradles. Treatment planning was performed using Pinnacle® 7.6, 8.0, or 9.0 (Philips Systems, Andover, MA) under the supervision of a radiation oncologist. All doses were calculated with correction for heterogeneous tissue density. For each patient, a diagnostic-quality CT scan acquired prior to RT and a treatment planning scan with an associated dose map were collected. Diagnostic scans were acquired using Philips Brilliance 16 ($n = 2$), 16P ($n = 10$), or 64 ($n = 8$) scanners and reconstructed using identical high-resolution lung filters and lung convolution kernels. Treatment planning scans were acquired using Philips Brilliance CT Big Bore ($n = 19$) or ACQSIM CT ($n = 1$) scanners and reconstructed using identical soft tissue filters and soft tissue convolution kernels. Dose maps were interpolated to match the voxel size in the planning scans using trilinear interpolation in CERR v. 4.0 Beta.[15] For all scans, semiautomated lung segmentation was performed by an experienced thoracic CT researcher (AC) using Pinnacle® 9.0 model-based segmentation. Two patients were later excluded from the study due to the inability to automatically identify distinctive landmark points in the patient's planning scan (see Sec. 2.B). Information on patient demographics, treatment, and image acquisition parameters for the remaining 18 patients is displayed in Table I.

2.B. Landmark selection and matching

Anatomically corresponding landmark points between scans were manually identified by three researchers (CC, BW, and JJ) to establish the reference correspondence between each patient's planning and diagnostic scans. To ensure that landmark points were identified across regions that represent a variety of dose levels, 2D 16×16-pixel regions of interest (ROIs) were first automatically placed within the lungs of the planning scan and distributed among four dose regions. Landmark points then were identified manually within each ROI in the planning scan and at the corresponding point in the diagnostic scan. The process to identify ROIs in the planning scan is displayed in Fig. 1. First, four "dose masks" were created using the dose map associated with the planning scan to identify regions in the lungs that lay within one of four dose regions (0–10 Gy, 10–30 Gy, 30–50 Gy, or >50 Gy). To focus analysis on normal lung tissue, pixels within the

TABLE I. Patient demographic information, treatment details, and image acquisition parameters.

Number of patients	18
	Male ($n = 8$)
	Female ($n = 10$)
Median patient age (range) (yr)	66 (49–79)
Number with smoking history	17
Lung cancer histology	NSCLC ($n = 13$)
	SCLC ($n = 5$)
Tumor lobe location	Upper ($n = 13$)
	Middle ($n = 2$)
	Lower ($n = 3$)
Treatment regimen	Concurrent chemo-RT ($n = 15$)
	Sequential chemo-RT ($n = 2$)
	RT only ($n = 1$)
Treatment planning strategy	3D-CRT ($n = 17$)
	IMRT ($n = 1$)
Median radiation dose (range) (Gy)	66 (60–70)
Dose per fraction (Gy)	2 ($n = 17$)
	2.5 ($n = 1$)
Median number of fractions (range)	33 (26–35)
Mean PTV volume (range) (cm^3)	555 (90–1234)
Median time between scan acquisition	Planning scans: 10 day (2–16)
and RT start (range)	Diagnostic scans: 19 day (4–75)
Treatment planning scan parameters	
Peak kilovoltage (kVp)	120
Slice thickness/spacing (mm)	3
Mean pixel spacing (range) (mm)	1.08 (0.94–1.28)
Dose grid size (mm)	$3 \times 3 \times 3$ ($n = 17$)
	$2 \times 2 \times 2$ ($n = 1$)
Motion management	Gated ($n = 13$)
	Free-breathing ($n = 5$)
Diagnostic scan parameters	
Peak kilovoltage (kVp)	120 ($n = 15$)
	140 ($n = 3$)
Slice thickness/spacing (mm)	1
Mean pixel spacing (range) (mm)	0.67 (0.57–0.83)

Note: NSCLC, Non-small cell lung cancer; SCLC, Small cell lung cancer; 3D-CRT, Three-dimensional conformal radiation therapy; IMRT, Intensity modulated radiation therapy; PTV, Planning target volume; RT, Radiation therapy.

boundaries of the planned target volume (PTV), which had been generated by a radiation oncologist during treatment planning using 5–10 mm margins, were excluded from the dose masks. Next, an automated program (iX v. 1.2.0.0)[16] was used to identify 100 high-pixel-value-gradient points in the planning scan. Stratified random sampling was used to ensure that points were uniformly distributed within each of the four dose masks. Two patients were excluded from the study due to either poor scan quality degraded by motion or the small volume of the lungs that received 30–50 Gy, preventing identification of 25 points in this dose region. Each of the 100 automatically identified points formed the center of a 2D 16×16-pixel ROI located entirely in one axial section of the planning scan, within which the user manually selected a single landmark point. The corresponding anatomically matched landmark point then was identified manually in the diagnostic scan. Placement of the matched landmark point in

the diagnostic scan was not constrained spatially, allowing the user to identify a landmark match at any point within any axial section of the diagnostic scan.

The graphical user interface for landmark identification and matching (Fig. 2) was developed using ABRAS v. 9.9b, an in-house software package for medical image visualization and manipulation.[17] Throughout the study, the user could pan/zoom, scroll through multiple CT sections, and alter the window and level. Due to differences in the breathing phase, resolution, and reconstruction methods between scans, it was not always possible to identify matched landmark points between scans within each of the automatically selected ROIs. All landmark points were reviewed by an experienced thoracic CT researcher (AC) and altered if necessary.

2.C. Observer training

Concurrence in landmark matching among the researchers identifying landmark points was tested using a training case, which was not used in the final analysis. The Euclidean distance between 100 matched landmark points identified by the experienced researcher (AC) and the three researchers (BW, CC, and JJ) was calculated. Multiple one-sided Wilcoxon signed rank tests were performed to determine if the distance was significantly larger than one-half of a voxel diagonal length ($r/2 = 1.7$ mm) in the planning scan. To maintain model significance of $\alpha = 0.05$, significance levels of individual tests were modified according to the Bonferroni method.

2.D. Automated landmark matching and dose calculation

Each patient's planning and diagnostic CT scans were coregistered using two deformable registration algorithms: Plastimatch demons registration[10] and Fraunhofer MEVIS's CT lung registration.[18] Registration proceeded as an iterative process at multiple resolution levels, with scans subsampled at progressively higher resolutions to allow for coarse image matching prior to alignment of small structures. During demons registration, a moving image is allowed to diffuse through a fixed image until matching is achieved, with the force and direction of motion at each iteration calculated using an optical flow equation. The Plastimatch implementation of this algorithm determined the number of iterations based on minimization of the mean-squared intensity difference between scans. The Fraunhofer MEVIS algorithm used a discrete optimization approach to determine an optimal mapping between CT scans. Specifically, a discretized objective function was minimized based on matching image gradients, while penalizing unrealistic volume changes, extreme deformation field curvature, and misaligned lung boundaries. For the Fraunhofer MEVIS algorithm, two sets of registration parameters were analyzed: "Fraunhofer MEVIS Fast" used three resolution levels for fast registration, while "Fraunhofer MEVIS EMPIRE10" was optimized for submission to the EMPIRE10 challenge[8] and used a four-level approach with no image subsampling at the finest level. For both Fraunhofer MEVIS methods, affine registration was performed
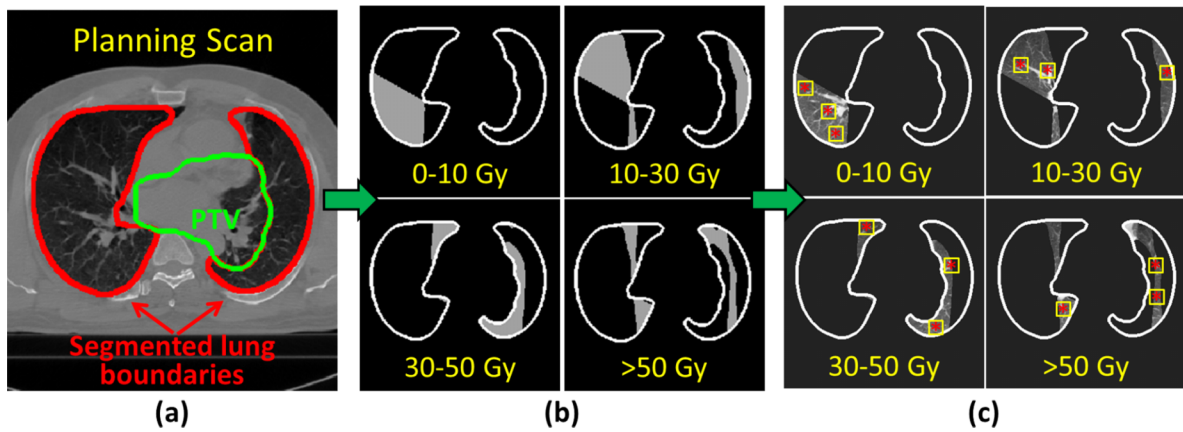
FIG. 1. Process for automatic identification of landmark points. (a) Planning CT scan with lung and PTV boundaries overlaid. (b) Dose masks generated using the dose map associated with the planning scan (lung boundaries excluding the PTV shown in white). (c) Points (*) that were automatically placed at high-pixel-value-gradient locations within each of the four dose masks formed the center of 16×16-pixel ROIs (yellow squares), within which landmark points were later selected manually.

prior to deformable registration. The parameters for registration with Plastimatch demons were selected based on past experience with this algorithm and parameter combination, which indicated highly accurate registration of diagnostic CT scans.[11] Registration with Plastimatch was performed during six stages, with image translation and affine registration occurring during the preliminary stages and demons registration at three resolution levels for the final three stages. Using the displacement vector fields (DVFs) output by each of the three registration algorithms, manually placed landmark points in the diagnostic scan (see Sec. 2.B) were automatically mapped to the corresponding planning scan (Fig. 3). For each algorithm, the Euclidean distance (d_E) between manually placed and automatically mapped landmark points in the planning scan was measured. Additionally, the difference in the planned dose between the two locations ($\Delta D = \mathrm{dose_{automatic}} - \mathrm{dose_{manual}}$) and the absolute dose difference ($|\Delta D|$) were calculated using the dose map associated with the planning

scan (Fig. 3). Regression modeling and analysis of variance (ANOVA) were used to determine whether significant differences in d_E existed among the three registration algorithms, accounting for patient-specific random effects. Furthermore, regression modeling was performed to characterize the individual contribution of d_E, dose (D), dose standard deviation ($\mathrm{SD_{dose}}$), and registration algorithm on $|\Delta D|$ according to the following formula:

$$|\Delta D|_{ijk} = \beta_0 + \mathrm{Patient}_i + \mathrm{Algorithm}_j + \beta_1 \times d_E + \beta_2 \times D + \beta_3 \times \mathrm{SD_{dose}} + \epsilon_{ijk}, \tag{1}$$

where β_0 is the intercept, $\mathrm{Patient}_i$ represents random patient-specific effects for each patient i, $\mathrm{Algorithm}_j$ represents how each of $j = 3$ registration methods (Plastimatch demons, Fraunhofer MEVIS Fast, or Fraunhofer MEVIS EMPIRE10) contributed to $|\Delta D|$, and β_1, β_2, and β_3 are the fitted slopes for d_E, D, and $\mathrm{SD_{dose}}$, respectively. ϵ_{ijk} is the residual error
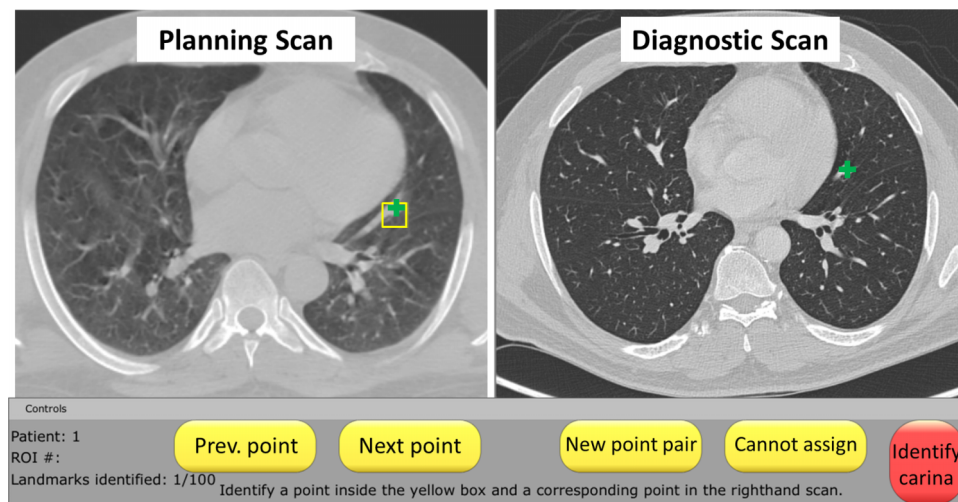

FIG. 2. Graphical user interface developed for landmark point identification and matching. Landmark points (green crosses) were manually placed within the ROI (yellow square), which had been automatically identified in the planning scan (left panel), and at the corresponding location in the diagnostic scan (right panel).
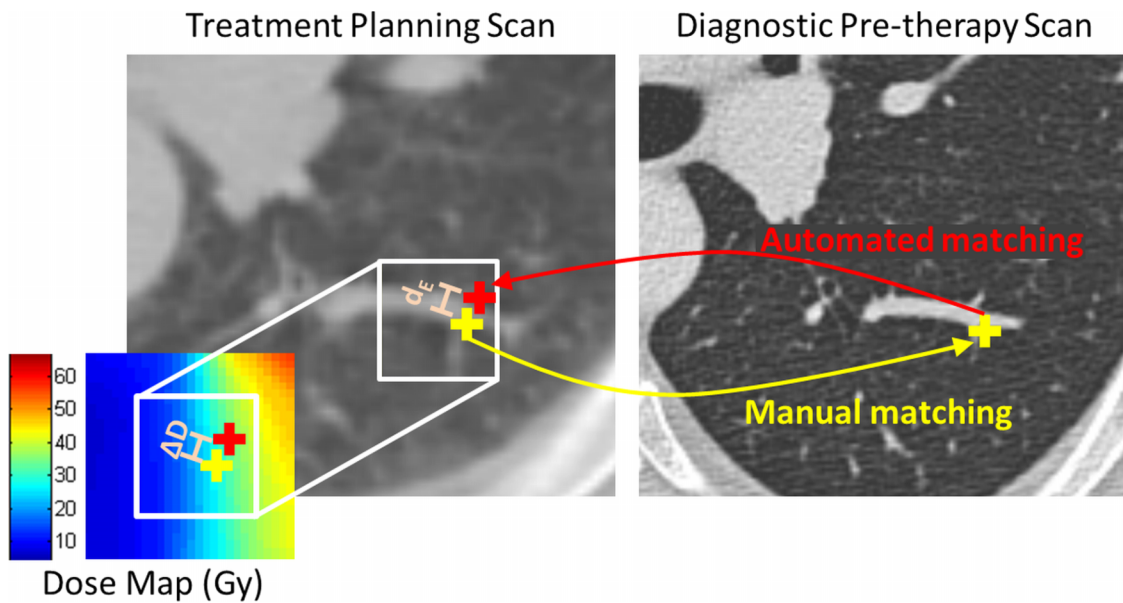
FIG. 3. Depiction of the method used to calculate the Euclidean distance (d_E) between manually and automatically placed landmark points and the resulting dose-mapping error (ΔD).

associated with the model. Values of D were calculated at each automatically identified landmark point in the planning scan dose map, and SD_{dose} was calculated in a 3×3 region centered at each of these points.

3. RESULTS

Comparable landmark matching accuracy was achieved among the multiple researchers using the training case (Table II). The distance between anatomically matched landmark points identified by different researchers in the diagnostic scans was not significantly larger than one-half of the size of a voxel in the planning scan.

For each of the 18 patients, between 58 and 93 landmark point pairs (median: 84) were identified for a total of 1472 point pairs, 16% of which were modified by the experienced CT researcher. For each patient, the mean values of d_E, ΔD, and $|\Delta D|$ were calculated for each registration algorithm. Table III displays the average of the 18 mean values of d_E, ΔD, and $|\Delta D|$ for each algorithm. Registration accuracy, as reflected by d_E, was similar for the Plastimatch demons and Fraunhofer MEVIS Fast algorithms but was greatly improved using Fraunhofer MEVIS EMPIRE10 (Fig. 4). For 15 of the 18 patients, Fraunhofer MEVIS EMPIRE10 achieved the lowest mean d_E. The fitted coefficients output using regression

modeling showed that, using Fraunhofer MEVIS EMPIRE10, d_E decreased significantly ($p < 0.05$) by 2 mm compared with the Plastimatch algorithm and by 2.4 mm compared with the Fraunhofer MEVIS Fast algorithm. This 0.4 mm difference in d_E between Fraunhofer MEVIS Fast and Plastimatch was also significant.

Despite CT scan alignment errors during deformable registration, the absolute dose-mapping error remained less than 2 Gy for the majority of points using each of the three algorithms (Fig. 5). Regression analysis and ANOVA indicated that d_E, D, SD_{dose}, and the registration algorithm significantly affected $|\Delta D|$ (Table IV). The model also indicated that, when d_E, D, and SD_{dose} were fixed, $|\Delta D|$ was increased by 0.12 Gy using Fraunhofer MEVIS EMPIRE10 and by 0.45 Gy using Fraunhofer MEVIS Fast compared with Plastimatch.

4. DISCUSSION

Despite registration error introduced using deformable registration, the subsequent effect of registration error on the dose-mapping error was small (≤ 4 Gy on average). The average registration error across patients was 7.4 mm using the Plastimatch demons registration algorithm and 7.6 and 5.2 mm using the Fraunhofer MEVIS Fast and EMPIRE10

TABLE II. Euclidean distance between landmark point matches identified by the researchers compared with the experienced thoracic CT researcher.

	Researcher 1	Researcher 2	Researcher 3
Median (IQR)	1.2 mm (0.7, 1.4)	1 mm (0.7, 1.4)	1 mm (0.7, 1.4)
Points with distance $\leq r/2$ (%)	82%	78%	75%

TABLE III. Average values of d_E, ΔD, and $|\Delta D|$ across patients for each of the three registration methods.

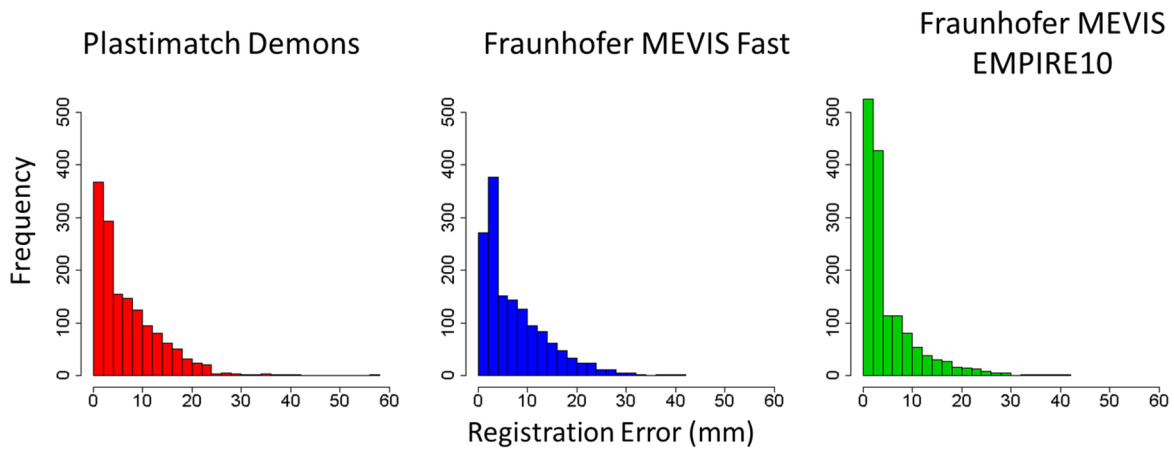|	Average d_E (mm) (range)	Average ΔD (Gy) (range)	Average $|\Delta D|$ (Gy) (range)
Plastimatch	7.4 (2.1–19.0)	−0.8 (−7.5–3.0)	3.5 (1.1–8.5)
MEVIS Fast	7.6 (2.3–15.9)	−1.4 (−6.9–1.7)	4.1 (1.2–10.6)
MEVIS EMPIRE10	5.2 (1.9–15.0)	−0.5 (−3.3–1.3)	2.8 (0.9–6.4)

FIG. 4. Distribution of registration errors (d_E) using each of the three registration algorithms.

algorithms, respectively (Table III). In a previous study in which healthy patient diagnostic CT scans were coregistered using Plastimatch demons, average registration error was only 0.39 mm.[11] The comparatively large registration errors measured in this study are likely due to differences in scan resolution, patient positioning, and respiratory phase between scans of a pair. Diagnostic scans were acquired during end-inspiration breath hold, while planning scans were acquired during free-breathing or end-exhalation. Scans were acquired using different scanners and reconstructed with different kernels. Additionally, the patient table used for the treatment planning scan was flat, while the diagnostic scanner table was curved, resulting in discrepancies in patient positioning. For this more complicated registration task, it is possible that use of a different set of registration parameters may decrease registration errors. Similar registration errors of 3–8 mm were measured in a study by Palma *et al.*[19] when diagnostic scans acquired 3 months following RT were coregistered with planning CT scans using B-splines deformable registration.

The average dose error across patients due to errors in registration was 3.5 Gy using Plastimatch demons, 4.1 Gy using Fraunhofer MEVIS Fast, and 2.8 Gy using Fraunhofer

MEVIS EMPIRE10 (Table III), which was roughly 4%–7% of the total prescribed dose. This average dose error is comparable to the 4% margin of error to the treatment site that is expected during treatment planning.[20] For the majority of points, dose-mapping error was below 2 Gy (approximately 3% of the prescribed dose). Linear regression indicated that a 1 mm increase in registration error would only result in a 0.42 Gy increase in dose-mapping error, on average. The small effect of registration error on dose-mapping error may be attributed to plan geometry. RT plans were generally designed so that 3–6 overlapping photon beams delivered a uniformly high dose to the tumor site and a uniformly low dose to the surrounding normal tissue, with a sharp dose gradient in between. Thus, large dose differences resulting from registration error are mainly to be expected in the relatively small volume of the lungs where a large dose gradient exists. Our findings are consistent with a study by Saleh-Sayah *et al.*,[13] who observed that, in low-dose-gradient regions, points displaced by up to 10 mm would result in an average dose-mapping error of <3.3 Gy for a single patient IMRT plan. In high-dose-gradient regions, however, the same average dose error would result from displacements as small as 1 mm. We observed that
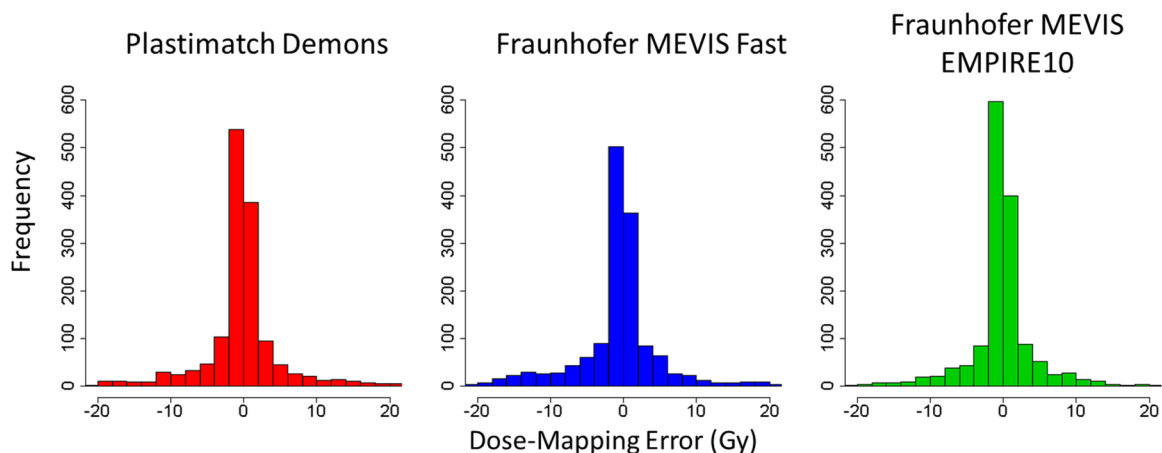

FIG. 5. Distribution (2 Gy/bin) of dose-mapping errors (ΔD) calculated using each of the three registration algorithms. Outliers, which comprised <3% of the data, are not displayed for visualization purposes.

TABLE IV. Model coefficients for Eq. (1) fitted using mixed effects regression modeling.

	Estimate (Gy)	Std. error (Gy)	F-statistic	p
β_0: Fitted model intercept (Algorithm$_0$ = PLASTIMATCH, $d_E = 0$, $D = 0$, SD$_{dose} = 0$)	−2.1	0.27	—	—
Algorithm$_1$: Change in $\|\Delta D\|$ using Fraunhofer MEVIS Fast	0.45	0.19	28	<0.001
Algorithm$_2$: Change in $\|\Delta D\|$ using Fraunhofer MEVIS EMPIRE10	0.12	0.19		
β_1: Change in $\|\Delta D\|$ per mm increase in d_E	0.42	0.01	740	<0.001
β_2: Change in $\|\Delta D\|$ per 1 Gy increase in D	0.05	0.004	318	<0.001
β_3: Change in $\|\Delta D\|$ per 1 Gy increase in SD$_{dose}$	1.4	0.11	199	<0.001

dose-mapping error was significantly affected by dose standard deviation (and thus, local dose gradient), a relationship also noted in other studies.[12,21] To our knowledge, our study is the first to measure dose-mapping errors in thoracic CT scans across multiple patients in a clinical database. Furthermore, by mapping dose between diagnostic and planning scans rather than between different phases of 4DCT scans or between fractions of RT treatment, our study allows for association of planned RT dose with radiation-induced toxicities that may develop outside of the dates of treatment delivery.

Regression modeling indicated that, when registration error, local dose gradient, and local dose were fixed, the Fraunhofer MEVIS Fast and EMPIRE10 algorithms increased $\|\Delta D\|$ by 0.45 and 0.12 Gy, respectively, relative to Plastimatch (Table IV). These results may appear to contradict the results in Table III, which show that dose-mapping error was smallest for Fraunhofer MEVIS EMPIRE10. Regression modeling indicated, however, that the improved dose accuracy observed using Fraunhofer MEVIS EMPIRE10 was due largely to improved registration accuracy. In other words, for points where d_E was the same between the Plastimatch and Fraunhofer MEVIS algorithms, ΔD was, on average, reduced using Plastimatch. DVFs for the Fraunhofer MEVIS algorithms demonstrated more pronounced curvature of the vector field (Fig. 6). A more curved DVF such as the one generated using Fraunhofer MEVIS is more likely to pass across multiple isodose levels, particularly when multiple impinging photon beams cause spatially linear dose gradients in 3D planning. As a result, the likelihood that points would be mapped from one dose region to another may be slightly increased using Fraunhofer MEVIS. While the superior registration accuracy achieved using the Fraunhofer MEVIS EMPIRE10 algorithm ultimately led to lower dose-mapping error on average (Table III), our findings indicate that dose alignment accuracy depends not only on the registration accuracy achieved by a deformable registration algorithm but also on the algorithm design itself. For instance, the curvature regularization term in the objective function used by the Fraunhofer MEVIS algorithm could be increased to allow for decreased DVF curvature. For the dose-mapping task, different considerations may be necessary to create a registration algorithm that is optimized to minimize dose-mapping errors.

Several potential weaknesses may impact the results of this study. First, manual identification of landmark point pairs may have resulted in incorrect identification of point pairs. Furthermore, because the landmark matching task was divided among three researchers, differences in the accuracy of landmark point placement may exist among researchers. These errors were reduced in several ways. First, observer training (Sec. 2.D) ensured that all researchers were familiar with the interface and that significant differences in landmark matching accuracy did not exist among the researchers for a training case. Second, all manually placed landmark points were reviewed by an experienced researcher and edited when necessary.

Due to the retrospective nature of database collection, variability existed among patients and treatment plans. PTV volume ranged from 90 to 1234 cm^3, which would impact treatment plan design to ensure appropriate tumor coverage. A single patient received intensity modulated RT (IMRT), while the remaining 17 patients were treated with three-dimensional conformal RT (3D-CRT). Though obvious differences in registration accuracy and dose-mapping error were not apparent for this single IMRT patient (average $d_E = 5.1$ mm and average $\|\Delta D\| = 2.2$ Gy), future studies involving a larger patient cohort should investigate the validity of our developed model for a variety of treatment planning strategies (e.g., IMRT, 3D-CRT, stereotactic body RT) and plan geometries, as different plan-
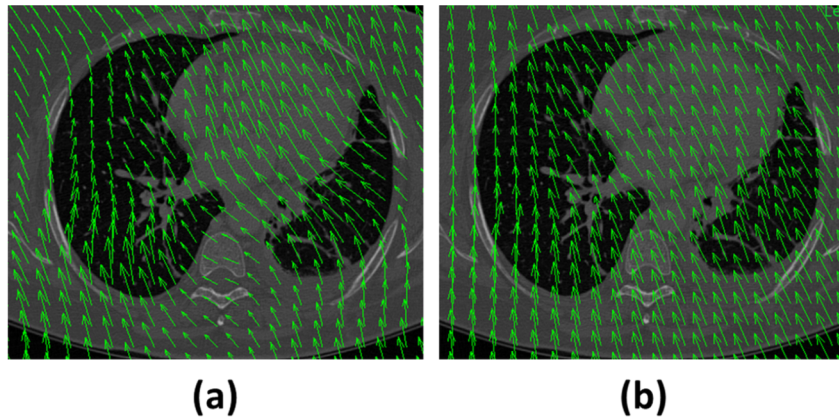
FIG. 6. (a) DVF generated using Fraunhofer MEVIS EMPIRE10 overlaid on a diagnostic scan section. The DVF for the Fast algorithm displayed similar curvature. (b) DVF generated using Plastimatch overlaid on the same diagnostic scan section. DVFs were visualized using VV (v. 1.3) (Ref. 22).

ning techniques are likely to affect the volume and location of the high-dose-gradient regions. The potential for application of our results to other tumor sites also remains to be investigated, as both organ motion and treatment plan design would vary for other sites.

In this study, dose-mapping error was modeled as a linear function of d_E, D, and SD_{dose}. During model development, we also considered quadratic rather than linear terms for the fixed effects and the inclusion of an interaction term between the registration algorithm and registration error (d_E). For all alternative models, the model fit was inferior to the original model [Eq. (1)], indicated by increases in the Akaike information criteria.

Dose-mapping error was measured throughout the entirety of the normal lung tissue and at multiple levels of radiation dose, allowing for future studies to be performed that focus on a variety of lung regions. For example, future studies could assess the uncertainty in dose to regions in close proximity to organs at risk. Alternatively, studies could correlate local radiation dose with regions displaying radiation-induced injury. Rather than calculate registration errors manually as was done in this study, several methods have been proposed to estimate registration error based on DVF analysis.[23–25] This concept was applied in a study by Salguero *et al.*,[12] who estimated dose-mapping error in a single patient 4D-CT planning scan based on both regional dose and the dispersion of the DVF during multiple deformations. To our knowledge, however, no study has been performed to measure the error that results when diagnostic scans are registered to planning scans, where dissimilarities between scans complicate the deformation task. Instead, the focus of past research has been to assess uncertainty in dose delivered during RT treatment either due to interfraction changes in patient positioning or breathing motion. The purpose of our study, however, was to facilitate association between radiation treatment and RT-induced changes that may only become apparent following treatment delivery (e.g., radiation pneumonitis or radiation fibrosis). Coregistration of diagnostic and planning scans would allow for direct spatial association between radiation dose and RT-induced toxicities, thus informing future patient care.

Our findings demonstrate that the accuracy of dose mapping through deformable registration is highly dependent on the local dose gradient and, to a lesser extent, the local dose. Thus, the dose alignment accuracy may be improved if, during the registration process, increased weight is given to the accurate registration of high-dose-gradient regions. Improvements in dose alignment accuracy may be especially important when performing coregistration of planning scans and diagnostic scans acquired after RT. Work in our lab demonstrated that, when diagnostic pre- and post-RT CT scans were coregistered, the presence of fibrotic changes in post-RT scans significantly decreased the registration accuracy by 2.2 mm using Plastimatch and by 1.2 mm using Fraunhover MEVIS Fast. In future studies when planning CT scans are coregistered with post-RT diagnostic scans, the effect of the presence of fibrotic changes on RT-induced damage would need to be considered. One way to improve dose alignment accuracy would be to employ landmark-guided registration,[26,27] with the majority of landmark points located in high-dose-gradient regions. The registration algorithm itself could instead be altered to more heavily weight high-dose-gradient regions when computing the similarity index between scans; unlike landmark-guided registration, this method has the advantage of being fully automated and thus more efficient for clinical implementation. Alternatively, post-RT diagnostic scans and planning CT scans both could be aligned with the pre-RT scan through deformable registration, allowing for association between post-RT diagnostic scans and planning scans without requiring direct coregistration of these two scans. This method also would have the advantage of allowing for treatment-related changes between pre- and post-RT scans to be assessed and associated with RT dose.

5. CONCLUSIONS

We measured the dose alignment accuracy when deformable registration was used to map planned dose between planning scans and diagnostic scans acquired before RT. Despite large inaccuracies in CT scan coregistration, dose-mapping

error remained less than 2 Gy for the majority of points using each of three registration algorithms. Dose-mapping error increased significantly as a function of registration error, local dose, local dose gradient, and the registration algorithm used. This methodology could be used in future research to characterize the dose-mapping error that is introduced when deformable registration is used to associate planned radiation dose with abnormalities in CT scans that develop following treatment with RT, potentially informing future patient care.

ACKNOWLEDGMENTS

[a]This work was presented, in part, at the 2014 AAPM Annual Meeting in Austin, TX.

[b]Author to whom correspondence should be addressed. Electronic mail: hal-hallaq@radonc.uchicago.edu; Telephone: 773-702-3309; Fax: 773-834-7299.

[1]J. Ma, J. Zhang, S. Zhou, J. L. Hubbs, R. J. Foltz, D. R. Hollis, K. L. Light, T. Z. Wong, C. R. Kelsey, and L. B. Marks, "Regional lung density changes after radiation therapy for tumors in and around thorax," Int. J. Radiat. Oncol., Biol., Phys. **76**, 116–122 (2010).

[2]A. Rovirosa and I. Valduvieco, "Radiation Pneumonitis," Clin. Pulm. Med. **17**, 218–222 (2010).

[3]L. B. Marks, S. M. Bentzen, J. O. Deasy, F. M. Kong, J. D. Bradley, I. S. Vogelius, I. El Naga, J. L. Hubbs, J. V. Lebesque, R. D. Timmerman, M. K. Martel, and A. Jackson, "Radiation dose-volume effects in the lung," Int. J. Radiat. Oncol., Biol., Phys. **76**, S70–S76 (2010).

[4]E. D. Yorke, A. Jackson, K. E. Rosenzweig, S. A. Merrick, E. S. Venkatraman, C. M. Burman, S. A. Liebel, and C. C. Ling, "Dose-volume factors contributing to the incidence of radiation pneumonitis in non-small-cell lung cancer patients treated with three-dimensional conformal radiation therapy," Int. J. Radiat. Oncol., Biol., Phys. **54**, 329–339 (2002).

[5]G. Rodrigues, M. Lock, D. D'Souza, E. Yu, and J. Van Dyk, "Prediction of radiation pneumonitis by dose-volume histogram parameters in lung cancer—A systematic review," Radiother. Oncol. **71**, 127–138 (2004).

[6]G. G. Zhang, T.-C. Huang, K. M. Forster, K.-P. Lin, C. Stevens, E. Harris, and T. Guerrero, "Dose mapping: Validation in 4D dosimetry with measurements and application in radiotherapy follow-up evaluation," Comput. Methods Programs Biomed. **90**, 25–37 (2008).

[7]E. C. J. Phernambucq, D. A. Palma, A. Vincent, E. F. Smit, and S. Senan, "Time and dose-related changes in radiological lung density after concurrent chemoradiotherapy for lung cancer," Lung Cancer **74**, 451–456 (2011).

[8]K. Murphy, B. van Ginneken, J. M. Reinhardt, S. Kabus, K. Ding, X. Deng, K. Cao, K. Du, G. E. Christensen, V. Garcia, T. Vercauteren, N. Ayache, O. Commowick, G. Malandain, B. Glocker, N. Paragios, N. Navab, V. Gorbunova, J. Sporring, M. de Bruijne, X. Han, M. P. Heinrich, J. A. Schnabel, M. Jenkinson, C. Lorenz, M. Modat, J. R. McClelland, S. Ourselin, S. E. Muenzing, M. A. Viergever, D. De Nigris, D. L. Collins, T. Arbel, M. Peroni, R. Li, G. C. Sharp, A. Schmidt-Richberg, J. Ehrhardt, R. Werner, D. Smeets, D. Loeckx, G. Song, N. Tustison, B. Avants, J. C. Gee, M. Staring, S. Klein, B. C. Stoel, M. Urschler, M. Werlberger, J. Vandemeulebroucke, S. Rit, D. Sarrut, and J. P. Pluim, "Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge," IEEE Trans. Med. Imaging **30**, 1901–1920 (2011).

[9]K. Brock, "Results of a multi-institution deformable registration accuracy study (MIDRAS)," Int. J. Radiat. Oncol., Biol., Phys. **76**, 583–596 (2010).

[10]G. Sharp, N. Kandasamy, H. Singh, and M. Folkert, "GPU-based streaming architectures for fast cone-beam CT image reconstruction and demons deformable registration," Phys. Med. Biol. **52**, 5771–5783 (2007).

[11]A. R. Cunliffe, H. A. Al-Hallaq, Z. E. Labby, C. A. Pelizzari, C. Straus, W. F. Sensakovic, M. Ludwig, and S. G. Armato, "Lung texture in serial thoracic CT scans: Assessment of change introduced by image registration," Med. Phys. **39**, 4679–4690 (2012).

[12]F. J. Salguero, N. K. Saleh-Sayah, C. Yan, and J. V. Siebers, "Estimation of three-dimensional intrinsic dosimetric uncertainties resulting from using deformable image registration for dose mapping," Med. Phys. **38**, 343–353 (2011).

[13]N. K. Saleh-Sayah, E. Weiss, F. J. Salguero, and J. V. Siebers, "A distance to dose difference tool for estimating the required spatial accuracy of a displacement vector field," Med. Phys. **38**, 2318–2323 (2011).

[14]A. Cunliffe, S. Armato, C. Straus, R. Malik, and H. Al-Hallaq, "A texture analysis approach to assess the severity of acute normal tissue changes in thoracic CT scans following radiation therapy for lung cancer," J. Thorac. Oncol. **8**, S969 (2013).

[15]J. O. Deasy, A. I. Blanco, and V. H. Clark, "CERR: A computational environment for radiotherapy research," Med. Phys. **30**, 979–985 (2003).

[16]K. Murphy, B. van Ginneken, S. Klein, M. Staring, B.J. de Hoop, M. A. Viergever, and J. P. W. Pluim, "Semi-automatic construction of reference standards for evaluation of image registration," Med. Img. Anal. **15**, 71–84 (2011).

[17]A. Starkey, W. Sensakovic, and S. Armato, "Abras: A portable application for observer studies and visualization," Int. J. Comput. Assist. Radiol. Surg. **6**, S193 (2011).

[18]J. Rühaak, S. Heldmann, T. Kipshagen, and B. Fischer, "Highly accurate fast lung CT registration," Proc. SPIE **8669**, 86690Y-1–86690Y-9 (2013).

[19]D. A. Palma, J. Van Sörnsen de Koste, W. F. A. R. Verbakel, and S. Senan, "A new approach to quantifying lung damage after stereotactic body radiation therapy," Acta Oncol. **50**, 509–517 (2011).

[20]B. J. Mijnheer, J. J. Battermann, and A. Wambersie, "What degree of accuracy is required and can be achieved in photon and neutron therapy?," Radiother. Oncol. **8**, 237–252 (1987).

[21]G. Janssens, J. Orban de Xivry, S. Fekkes, A. Dekker, B. Macq, P. Lambin, and W. van Elmpt, "Evaluation of nonrigid registration models for interfraction dose accumulation in radiotherapy," Med. Phys. **36**, 4268–4276 (2009).

[22]P. Seroul and D. Sarrut, "VV: A viewer for the evaluation of 4D image registration," MIDAS Journal—Systems and architectures for computer Assisted Interventions (MICCAI 2008 Workshop) (available online: http://hdl.handle.net/10380/1458) (2008).

[23]M. Hub and C. P. Karger, "Estimation of the uncertainty of elastic image registration with the demons algorithm," Phys. Med. Biol. **58**, 3023–3036 (2013).

[24]C. Vaman, D. Staub, J. Williamson, and M. J. Murphy, "A method to map errors in the deformable registration of 4DCT images," Med. Phys. **37**, 5765–5776 (2010).

[25]H. Zhong, T. Peters, and J. V. Siebers, "FEM-based evaluation of deformable image registration for radiation therapy," Phys. Med. Biol. **52**, 4721–4738 (2007).

[26]H. J. Johnson and G. E. Christensen, "Consistent landmark and intensity-based image registration," IEEE Trans. Med. Imaging **21**, 450–461 (2002).

[27]H. Lu, P. C. Cattin, and M. Reyes, "A hybrid multimodal non-rigid registration of MR images based on diffeomorphic demons," in *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE, Buenos Aires, Argentina, 2010), pp. 5951–5954.