Radiology

Quality Assurance Assessment of Diagnostic and Radiation Therapy–Simulation CT Image Registration for Head and Neck Radiation Therapy: Anatomic Region of Interest–based Comparison of Rigid and Deformable Algorithms[1]

Abdallah S. R. Mohamed, MD, MSc[2]
Manee-Naad Ruangskul, MD[3]
Musaddiq J. Awan, MD[4]
Charles A. Baron, BA[5]
Jayashree Kalpathy-Cramer, PhD
Richard Castillo, PhD
Edward Castillo, PhD
Thomas M. Guerrero, MD, PhD
Esengul Kocak-Uzel, MD[6]
Jinzhong Yang, PhD
Laurence E. Court, PhD
Michael E. Kantor, BS
G. Brandon Gunn, MD
Rivka R. Colen, MD
Steven J. Frank, MD
Adam S. Garden, MD
David I. Rosenthal, MD
Clifton D. Fuller, MD, PhD

Purpose: To develop a quality assurance (QA) workflow by using a robust, curated, manually segmented anatomic region-of-interest (ROI) library as a benchmark for quantitative assessment of different image registration techniques used for head and neck radiation therapy–simulation computed tomography (CT) with diagnostic CT coregistration.

Materials and Methods: Radiation therapy–simulation CT images and diagnostic CT images in 20 patients with head and neck squamous cell carcinoma treated with curative-intent intensity-modulated radiation therapy between August 2011 and May 2012 were retrospectively retrieved with institutional review board approval. Sixty-eight reference anatomic ROIs with gross tumor and nodal targets were then manually contoured on images from each examination. Diagnostic CT images were registered with simulation CT images rigidly and by using four deformable image registration (DIR) algorithms: atlas based, B-spline, demons, and optical flow. The resultant deformed ROIs were compared with manually contoured reference ROIs by using similarity coefficient metrics (ie, Dice similarity coefficient) and surface distance metrics (ie, 95% maximum Hausdorff distance). The nonparametric Steel test with control was used to compare different DIR algorithms with rigid image registration (RIR) by using the post hoc Wilcoxon signed-rank test for stratified metric comparison.

Results: A total of 2720 anatomic and 50 tumor and nodal ROIs were delineated. All DIR algorithms showed improved performance over RIR for anatomic and target ROI conformance, as shown for most comparison metrics (Steel test, $P < .008$ after Bonferroni correction). The performance of different algorithms varied substantially with stratification by specific anatomic structures or category and simulation CT section thickness.

Conclusion: Development of a formal ROI-based QA workflow for registration assessment demonstrated improved performance with DIR techniques over RIR. After QA, DIR implementation should be the standard for head and neck diagnostic CT and simulation CT allineation, especially for target delineation.

© RSNA, 2014

Online supplemental material is available for this article.

Radiology

Deformable image registration (DIR) is an increasingly common tool for applications in image-guided radiation therapy (1–3). DIR, as a tool for motion assessment and correction in tumors that move with respiration, as well as adaptive recontouring of target or anatomic volumes that alter over time, is becoming more widely used as vendors integrate DIR solutions into commercial software packages (4–9). Additionally, emerging data suggest that for head and neck cancers, DIR has demonstrable technical performance gain compared with rigid image registration (RIR) for adaptive radiation therapy, wherein a simulation computed tomographic (CT) data set is coregistered with CT-on-rails or cone-beam CT data acquired during the course of radiation treatment (10–13).

Three-dimensional radiation therapy–simulation CT (hereafter, simulation CT) data sets are the initial component of radiation planning. Simulation CT data sets are then manually segmented to define both tumor and normal tissue volumes, with subsequent dose calculation performed by using voxel-based electron density maps. Consequently, as the key imaging step in radiation therapy, all subsequent patient treatment dose delivery is entirely dependent on the quality of simulation CT processes (eg, target delineation, organ-at-risk segmentation, beam and intensity optimization, and dose calculation) (14).

The intramodality registration of pre- and posttherapy head and neck diagnostic CT with simulation CT data are valuable in multiple radiation therapy applications, such as target delineation and mapping sites of posttherapy local-regional recurrences with the original simulation CT images and dose grid (15–18). Nevertheless, such intramodality fusion of diagnostic CT images and simulation CT images is less commonly described in the literature and presents specific obstacles to image registration. First, diagnostic CT acquisition is routinely performed by using a curved tabletop without standardized head positioning, while the simulation CT images are obtained by using a custom thermoplastic immobilization mask on a flat-topped table, resulting in positional differences of head and neck tissues (15,16). In several head and neck cancers, institutional use of an intraoral immobilization and displacement device, such as a custom dental stent (19), results in placement of a new structure in the simulation CT examination that was not present in the diagnostic CT examination. Additionally, diagnostic CT typically entails the use of intravenous contrast material for tumor assessment, while at our facility and many others, intravenous contrast material is not used for simulation CT, resulting in intensity differentials for the same structures (10). Acquisition parameters (eg, section thickness reconstruction [STR], field of view, and peak voltage) may not be standardized between diagnostic CT and simulation CT. Finally, in

many instances, owing to either tumor progression or intervening therapy (surgical resection or induction chemotherapy), the anatomy is fundamentally altered between diagnostic CT and simulation CT. These factors, among others, make registration of diagnostic CT images with simulation CT images a nontrivial task.

As part of larger efforts to improve head and neck target delineation, as well as defining spatially accurate mapping of local-regional failure sites, the purpose of this study was to develop a quality assurance (QA) workflow by using a robust, curated, manually segmented anatomic region-of-intereset (ROI) library as a benchmark for quantitative assessment of different image registration techniques used for coregistration of head and neck simulation CT images with diagnostic CT images.

Advance in Knowledge

■ Most overlap and surface distance metrics investigated showed a detectable improvement in conformance with a manual region of interest (ROI) comparator by using deformable image registration (DIR) techniques over rigid image registration (RIR) for all pooled ROIs (Steel test, $P < .008$ after Bonferroni correction for multiple comparisons) in the specific setting of registration of diagnostic CT images with radiation therapy–simulation CT images.

Implications for Patient Care

■ After proper quality assurance, DIR should replace RIR for radiation therapy applications involving registration of head and neck diagnostic CT images with radiation therapy–simulation CT images.

■ Validation of any image registration strategy is crucial before implementation in specific clinical scenarios.

Materials and Methods

Study Population

Simulation CT and diagnostic CT Digital Imaging and Communications in Medicine (DICOM) files of 22 patients with head and neck cancer treated at our institution between August 2011 and May 2012 were retrospectively retrieved with institutional review board approval. Inclusion criteria were pathologically proven diagnosis of squamous cell carcinoma of the head and neck, treatment with curative-intent intensity-modulated radiation therapy, and availability of nonenhanced simulation CT and contrast material–enhanced diagnostic CT images for each patient, with a maximum time interval of 4 weeks between the two examinations to minimize errors attributed to anatomic changes associated with radiation therapy or disease progression. Twenty patients were eligible, while two patients were excluded—one for having massive disease progression and the other for undergoing surgical resection during the interval between diagnostic CT and simulation CT. Patients and treatment characteristics are summarized in Table 1.

Imaging Characteristics

Nonenhanced simulation CT images were acquired after immobilizing patients with thermoplastic head and neck shoulder masks, with a section thickness of 1–3.75 mm and an x-ray tube current of 100–297 mA at 120 kVp. Display field of view was 500 mm; axial images were acquired by using a matrix of 512 × 512 pixels and reconstructed with a pixel size of 0.98 × 0.98 mm along the x- and y-axis. Comparatively, contrast-enhanced diagnostic CT images were acquired with a section thickness of 1.25–3.75 mm and an x-ray tube current of 160–436 mA at 120 kVp. Display field of view was 236–300 mm, and axial images were acquired by using a matrix of 512 × 512 pixels, reconstructed with a pixel size of 0.46 × 0.46 mm to 0.59 × 0.59 mm along the x- and y-axis. One hundred twenty milliliters

of contrast material were injected at a rate of 3 mL/sec, followed by scanning after a 90-second delay. (Detailed acquisition parameters are given in Table E1 [online].)

Manual Segmentation of Reference Anatomic ROIs

For each patient, a series of 68 reference anatomic ROIs (18 bone, three cartilaginous, seven glandular, 30 muscular, six soft-tissue, and four vascular ROIs), in addition to gross primary tumor volume and gross nodal target, were manually contoured on every patient's diagnostic CT images and simulation CT images by both a 3rd-year resident physician observer (M.N.R.) and a medical student (C.A.B.). Contours were subsequently approved on a daily basis by a radiation oncologist with 7 years of experience (A.S.R.M.), and, finally, ROIs were reviewed by an expert attending head and neck radiation oncologist with 8 years of experience (C.D.F.). Manual segmentation was performed by using commercial treatment planning software (Pinnacle 9.0; Phillips Medical Systems, Andover, Mass). (Details of the segmented ROIs are available as Figure E1 and Table E2 [online].)

Image Registration

For each patient's DICOM data set, image registration was performed by using baseline RIR, allowing automatic scalable rigid registration of the diagnostic CT images with the simulation CT images by using a block-matching in-house graphics processing unit–based algorithm. Subsequently, each data set was registered by using a series of open-access and commercial registration algorithms. Two kinds of commercial multistep DIR software were examined: atlas based (CMS ABAS 0.64; Elekta, Stockholm, Sweden) and B-spline (VelocityAI 2.8.1; Velocity Medical, Atlanta, Ga). Atlas-based DIR, presented previously (20), consisted of five registration steps to deform the original diagnostic CT image to the simulation CT image: linear registration, head pose correction,

poly-smooth nonlinear registration, dense mutual-information deformable registration, and final refinement by using a deformable surface model (20). The B-spline method consisted of three registration steps: manually adjusted rigid edit, automatic rigid registration, and automatic deformable registration (21).

Table 1	
Patient Characteristics	
Characteristic	Value
Median age (y)	63 (50–78)
Patient sex	
No. of men	15
No. of women	5
T stage	
T1	4
T2	5
T3	8
T4	2
Tx	1
N stage	
N0	3
N1	4
N2	11
N3	1
Nx	1
Primary sites	
Base of tongue	7
Tonsil	6
Oral cavity	1
Larynx	3
Nasopharyngeal or maxillary sinus	2
Salivary gland	1
Treatment	
IMRT alone	1
Concurrent chemotherapy and IMRT	5
Induction chemotherapy followed by concurrent chemotherapy and IMRT	9
Induction chemotherapy followed by IMRT	3
Surgery followed by postoperative chemotherapy and IMRT	1
Induction chemotherapy and IMRT followed by surgery	1

Note.—Data are numbers of patients, unless indicated otherwise. Numbers in parentheses are the range. IMRT = intensity-modulated radiation therapy.

Radiology

Likewise, two noncommercial software algorithms, ITK demons (Kitware, Clifton Park, NY) (22) and optical flow (Castillo, Houston, Tex) (23), were investigated. Noncommercial algorithms are often dependent on numerous manually entered variable parameters that determine the effectiveness of an algorithm in registration. To provide valid comparison of commercial and noncommercial algorithms, an optimization step was performed by using approximately 100 human-verified identical landmark points on both diagnostic CT and simulation CT images and the methods that we have described previously (24). With these methods, parameters were iteratively varied to create varying deformation fields to map the diagnostic CT images with the simulation CT images. Subsequently, these deformation fields were applied to the landmark points on the diagnostic CT images to obtain deformed landmark points on the simulation CT images. Euclidean distances in corresponding landmark points between the deformed landmark points and the actual landmark points on the simulation CT images were calculated. Parameters that minimized the total Euclidean point distances obtained from applying these noncommercial algorithms to these landmark points were used in the ROI comparison described, as follows.

Deformation vector fields were obtained from each image registration algorithm to map the deformation of the diagnostic CT images onto the simulation CT images. For the commercial algorithms, the deformation transformed definitive voxels from the original diagnostic CT images onto the simulation CT images, while for the noncommercial algorithms, the deformation field was from the rigid CT images. Subsequently, in a custom-written MATLAB program (MATLAB R2012a; MathWorks, Natick, Mass), these deformation fields were applied to ROIs segmented on the diagnostic CT images to convert them into "deformed ROIs" on the simulation CT images. Calculations (described as follows) were performed by comparing
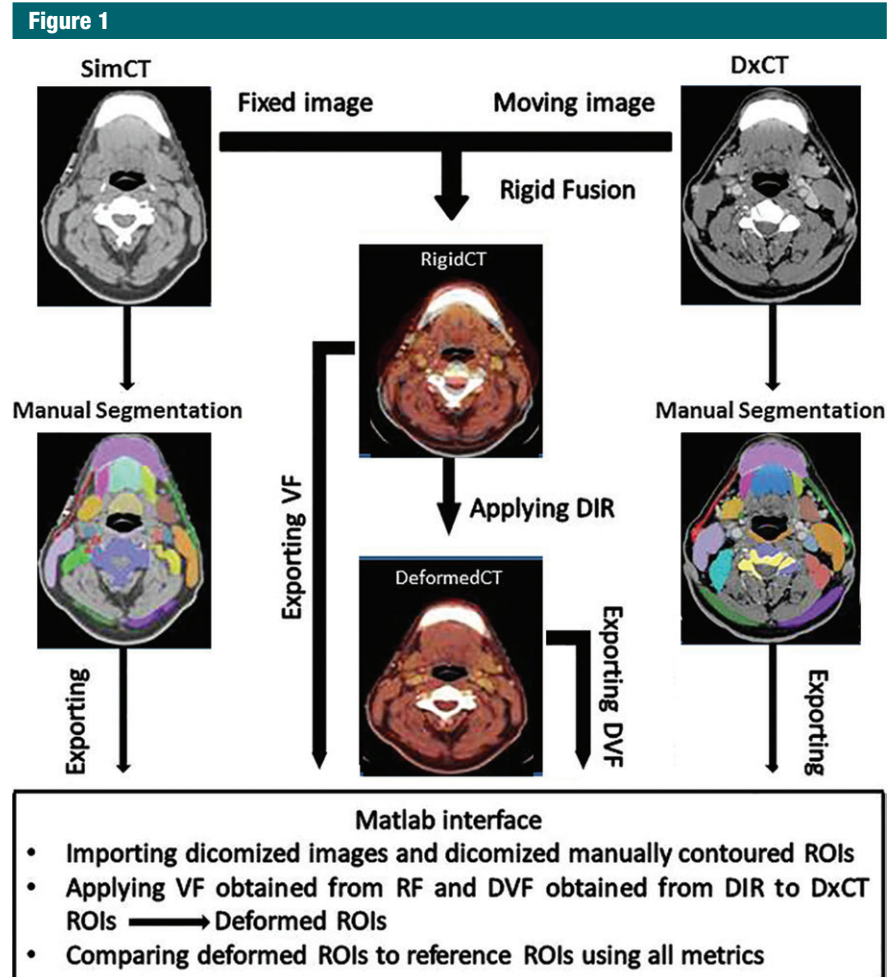

Figure 1: Schematic illustration demonstrates the QA workflow process. *DVF* = deformation vector fields, *DxCT* = diagnostic CT, *RF* = rigid fusion, *SimCT* = simulation CT, *VF* = vector fields.

the "deformed ROIs" with the human-segmented ROIs on the simulation CT images.

Registration Algorithm Assessment

After completion of RIR and DIR of native DICOM image files, the resultant deformed ROIs were compared quantitatively with manually contoured reference ROIs for each of the 68 anatomic structures listed, plus tumor and nodal targets. Figure 1 shows a schematic illustration of the QA workflow process developed on the basis of a previously presented software resource (TaCTICS [Target Contour Testing/Instructional Computer Software], *https://github.com/kalpathy/tacticsRT*) (25–27).

For each registration-deformed ROI and reference ROI pair, the following ROI-based overlap metrics (28–30) were assessed: volume overlap, maximum and 95% maximum Hausdorff distance, and false-positive, false-negative, and standard Dice similarity coefficients (DSCs).

Individual metrics are detailed in Table 2. The different registration algorithms were compared by using different metrics for pooled ROI overlap, then compared after stratification by individual ROI, anatomic subgroup, and STR.

Interobserver Variability

To interrogate the effect of interobserver variability in manual ROI

Table 2

Description of Used Metrics

Metric	Symbolic Expression	Description	Full Agreement
Volume overlap	$VO = \dfrac{|A \cap R|}{|R|}$	The portion of the reference ROI that is overlapped by the segmentation	1
DSC	$DSC = \dfrac{2|A \cap R|}{|A| + |R|}$	The portion of overlap between the reference ROI and segmentation relative to the size of the reference ROI plus the size of the segmentation	1
False-negative Dice	$FND = \dfrac{2|\overline{A} \cap R|}{|A| + |R|}$	The volume that the segmentation missed of the reference ROI relative to the size of the reference ROI plus the size of the segmentation	0
False-positive Dice	$FPD = \dfrac{2|A \cap \overline{R}|}{|A| + R}$	The volume of the segmentation not found within the reference ROI relative to the size of the reference ROI plus the size of the segmentation	0
95% Maximum Hausdorff distance	$H(A,R) = \max[h(A,R), h(R,A)]$ ∍ $h(A,R) = \max_{a \in A}[d(A,R)]$ & $d(A,R) = \min_{r \in R}\|a - r\|$	The maximum distance between a point in the segmentation and that of the reference ROI; for 95% maximum Hausdorff, ±5 percentile outliers are discarded	…

Note.—a = a point in A, A = algorithm-dependent deformed ROI volume, $d(A,R)$ = minimum surface distance between each point in A from the surface R, FND = false-negative Dice, FPD = false-positive Dice, $h(A,R)$ = the maximum of the minimum of the surface distances between every point in surface A and the surface R, $H(A,R)$ = Hausdorff distance between A and R, r = a point in R, R = reference anatomic manually segmented ROI volume, VO = volume overlap.

delineation on the output metrics of different registration techniques post hoc, a subset of 12 ROIs (two ROIs per each anatomic subgroup) were delineated by three expert observers in clinical target delineation (A.S.R.M., with 7 years of experience; a nonauthor, with 8 years of experience; and E.K.U., with 5 years of experience) on paired simulation CT and diagnostic CT sets in two patients. The same workflow methods described earlier were used to compare overlap and surface distance metrics obtained from expert segmentation with those of the primary observer for each registration technique.

Statistical Analysis

Statistical assessment was performed by using JMP version 10.2 software (SAS institute, Cary, NC). To assess algorithm performance for each metric listed, distributional statistics for listed metrics (Table 2) were tabulated for each anatomic ROI.

To determine the relative degree of potential difference of DIR compared with RIR, overlap metrics for anatomic and target ROIs mapped by using DIR were compared with those mapped rigidly. P value thresholding for multiple comparisons with Bonferroni correction was conducted by dividing the requisite α threshold of .05 by the number of subset comparisons, with the resultant P value specified for each comparison (vide infra). Distributional differences in DIR algorithms (atlas based, demons, optical flow, and B-spline) as compared with RIR alone were performed by using the nonparametric Steel test with control (31) for between-algorithm differences, with RIR designated as the standardized comparator (control). Between-group comparisons were performed for all metrics for pooled ROIs and for each ROI separately and were reported with P value thresholding for multiple comparisons applied graphically.

Additional post hoc comparison of metrics after stratification according to STR and ROI anatomic subgroup (bones, cartilage, muscles, glands, vessels, and soft tissues) was performed, with comparison by using Wilcoxon signed-rank (32) (paired comparison) or Kruskal-Wallis (33) (comparison of three or more groups) nonparametric tests when comparing across strata with P value thresholding for multiple comparisons. For evaluation of interobserver dependency, the Cronbach α method was performed to assess agreement between expert observers for all delineated ROIs for all tested registration methods for DSC metrics.

Results

Algorithm Comparative Performance

A total of 2720 anatomic ROIs were delineated (68 per each DICOM file) for all included patients' paired simulation

Figure 2: Heat map illustrates the relative performance of each of the DIR algorithms over the RIR for each anatomic structure by using all comparison metrics. Better performance is indicated by colors that appear toward the blue end of the color scale. P value thresholding for multiple comparisons was used with the first threshold at P less than .008 for correction of multiple comparisons across four distinct registration algorithms (α = .05/6 "pairwise comparison of four DIR algorithms") and P less than .0001 for multiple comparisons across 68 distinct ROIs. Values shaded solid red are nonsignificantly different from rigid segmentation. FN-DSC = false-negative DSC, FP-DSC = false-positive DSC, GTV-N = gross nodal volume, GTV-P = gross primary tumor volume, MPC = middle pharyngeal constrictor, OF = optical flow, SPC = superior pharyngeal constrictor, ST = soft tissues, VO = volume overlap, 95% HD = 95% Hausdorff distance.

Figure 2

Significance of Steel test with control (Each method compared to RIR)
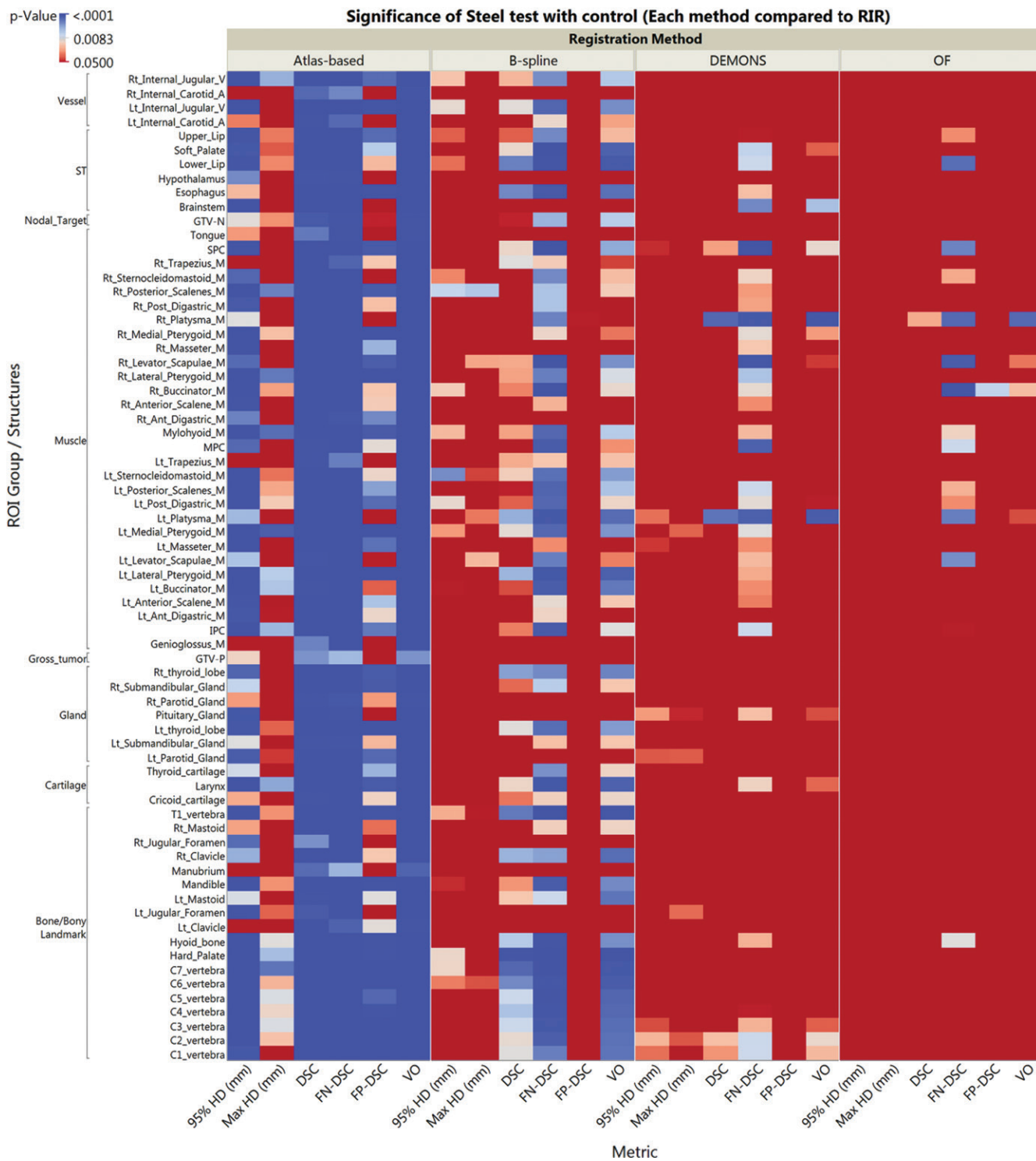
CT and diagnostic CT sets, and a total of 50 tumor and nodal ROIs were delineated in 15 of 20 patients' paired sets (five patients had no radiologic gross tumor or nodal targets after induction chemotherapy and/or surgery). Excepting optical flow for surface distance metrics (ie, 95% and maximum Hausdorff distance) and demons for false-negative Dice, for all investigated DIR algorithms there was a detectable improvement in conformance with a manual ROI comparator over RIR for all pooled ROIs by using all comparative metrics (Steel test, $P <$ .008 after Bonferroni correction for multiple comparisons; a P value heat map is available as Figure E2 [online]). Estimation of differences between algorithms varied substantially with stratification by specific ROI; magnitude of effect size for the difference as compared with RIR control is illustrated as a P value heat map with color thresholding for multiple comparison correction (Fig 2). Figure 3 shows an example of the visual comparison of the overlap between deformed ROIs and reference ROIs by using RIR versus DIR.

For similarity coefficient metrics, the tested atlas-based algorithm appears to have the best performance in this specific head and neck application (median DSC, 0.68; median false-negative Dice coefficient, 0.06; and median false-positive Dice coefficient, 0.5). Likewise, for surface distance metrics, the atlas-based algorithm had the least median distance error (4.6-mm 95% Hausdorff distance; 10.6-mm maximum Hausdorff distance). (Figures E3 and E4 [online] further illustrate the overlap and surface distance metrics used, respectively, for all ROIs for each registration method as a graphic table.)

Algorithm Performance according to Anatomic Subgroup

The performance of each registration method varied significantly across different ROI subgroups (bones, cartilage, muscles, glands, vessels, and soft tissues) as assessed for different metrics by using the Kruskal-Wallis test ($P <$ .05), excepting RIR (for volume overlap and 95% Hausdorff distance), B-spline (for volume overlap), and
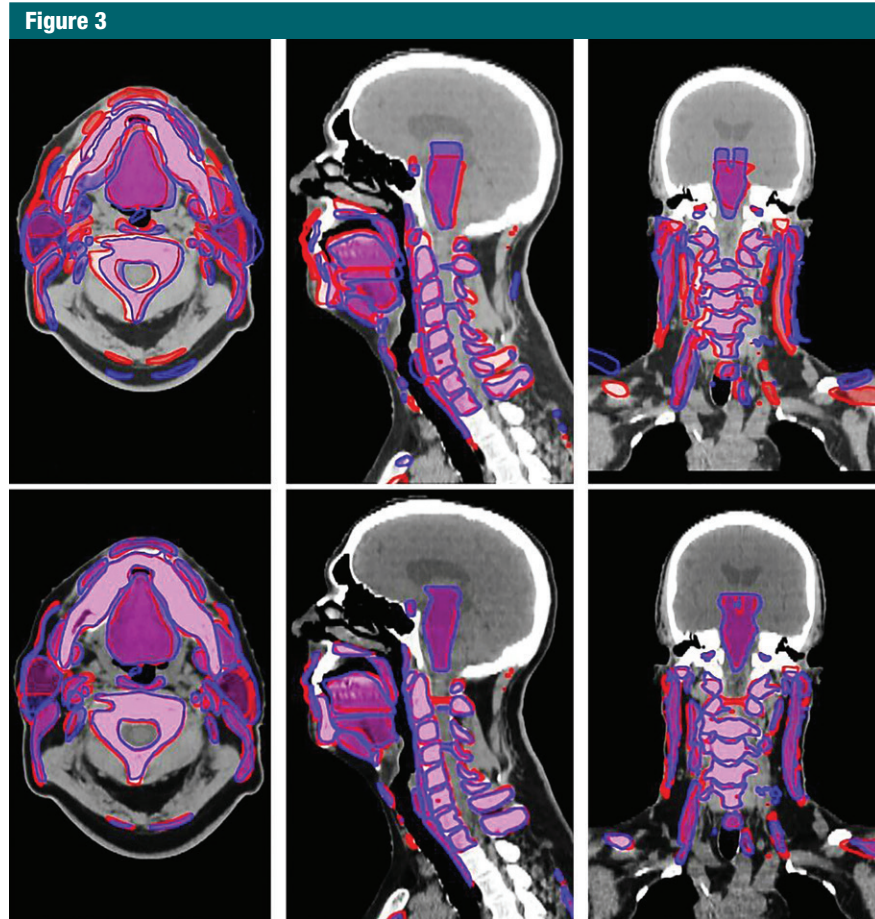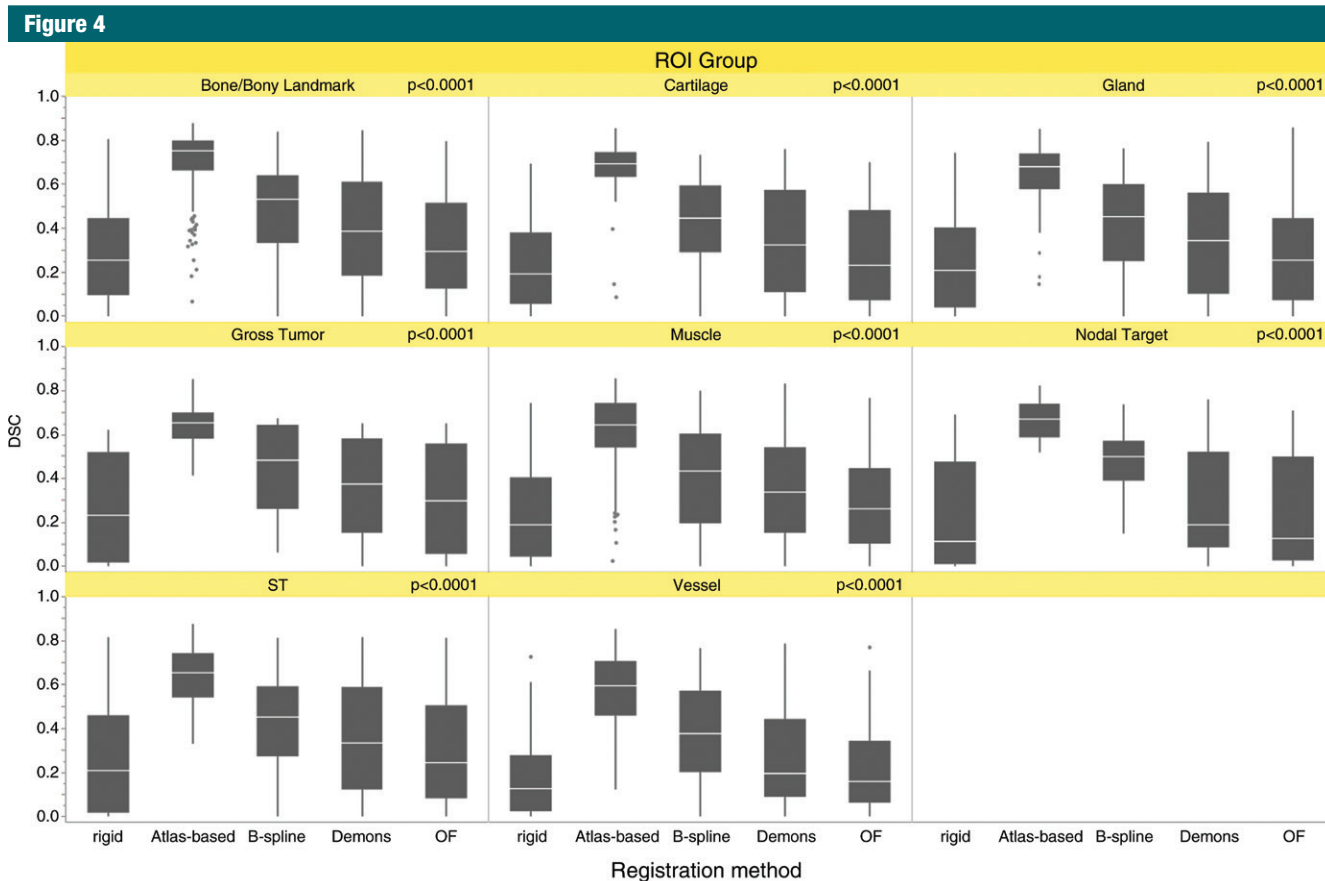

Figure 3

Figure 3: Images demonstrate the deformed-reference ROI overlay. The top panels show examples of deformed ROIs in blue and reference ROIs in red by using RIR, while the lower panels show the same ROIs when registered by using DIR.

demons (for 95% Hausdorff distance). To directly compare specific anatomic subgroup performance for each registration algorithm, further analysis with the paired Wilcoxon test showed that for most registration algorithms, bone and cartilaginous ROIs were significantly more concordant than muscular and vascular ROIs (Bonferroni-corrected $P <$ .003 for multiple comparisons across the six ROI subgroups). Likewise, for each anatomic category, ROI conformance varied significantly across different registration methods for all assessed metrics (Kruskal-Wallis test, $P <$ .05) with comparatively better performance of atlas-based DIR, followed by B-spline, in all ROI subgroups (Bonferroni-corrected $P <$ .005 for multiple comparisons across

the five registration methods), as illustrated visually in Figure 4 for DSC and 95% Hausdorff distance metrics. The tumor and nodal ROI conformance was best achieved with atlas-based software with gross primary tumor volume median DSC of 0.65 and 95% Hausdorff distance of 4.6 mm and gross nodal volume median DSC of 0.67 and 95% Hausdorff distance of 4.5 mm, respectively.

Algorithm Performance according to Section Thickness Reconstruction

STR of diagnostic CT for all but two cases was uniform at 1.25 mm, while the simulation CT STR was variable, with 10 cases reconstructed at a section thickness of less than 3 mm (1–2.5 mm) and 10 cases reconstructed at a

Figure 4

a.

Figure 4: (a) Box plots of DSC analysis for each registration method are shown according to anatomic ROI category. The pale line within the box indicates the median value, while the box limits indicate the 25th and 75th percentiles. The lines represent the 10th and 90th percentiles, and the dots represent outliers *(Fig 4 continues).*

section thickness of 3 mm or more (3–3.75 mm). The effect of variability in simulation CT STR on the performance of individual registration methods was most evident for the tested B-spline algorithm as shown in Table 3, a system dependency not typically considered in image-registration QA but of practical importance in clinical image acquisition.

Interobserver Variability

In post hoc interobserver dependency assessment, for each registration technique there was no significant (all $P >$ .05) difference between the mean values for all metrics of all expert observers when compared with the primary observer, except for the 95% Hausdorff distance for observer 3 in atlas-based and B-spline algorithms. Cronbach α

assessment showed a minimum α value of 0.6527 (acceptable interobserver agreement) for RIR and a maximum α value of 0.88 for B-spline (good interobserver agreement). This confirms that the interobserver manual delineation variability was unlikely to substantively affect the outcomes of registration in this study (Fig E5 [online]).

Discussion

Our results confirm that the tested DIR algorithms provided detectable performance advantages over RIR in our specific head and neck diagnostic CT and simulation CT data set for most ROI-based metrics. The performance of specific DIR algorithms varied across anatomic ROIs, with greater

conformance of registering bone and cartilaginous ROIs with reference ROIs than that for muscle and vascular ROIs. Furthermore, the difference between the registration accuracy of different structures of the anatomic class varied substantially. For example, certain bone structures (eg, the clavicles) showed more distance error and reduced conformance as compared with other bone ROIs, likely due to the wide variation in shoulder position between both images. Another example is the relatively higher number of registration errors of tongue musculature ROIs and velar ROIs in the region adjacent to intraoral dental stents with simulation CT. Our QA method also demonstrated unanticipated algorithm dependencies, such as STR, between
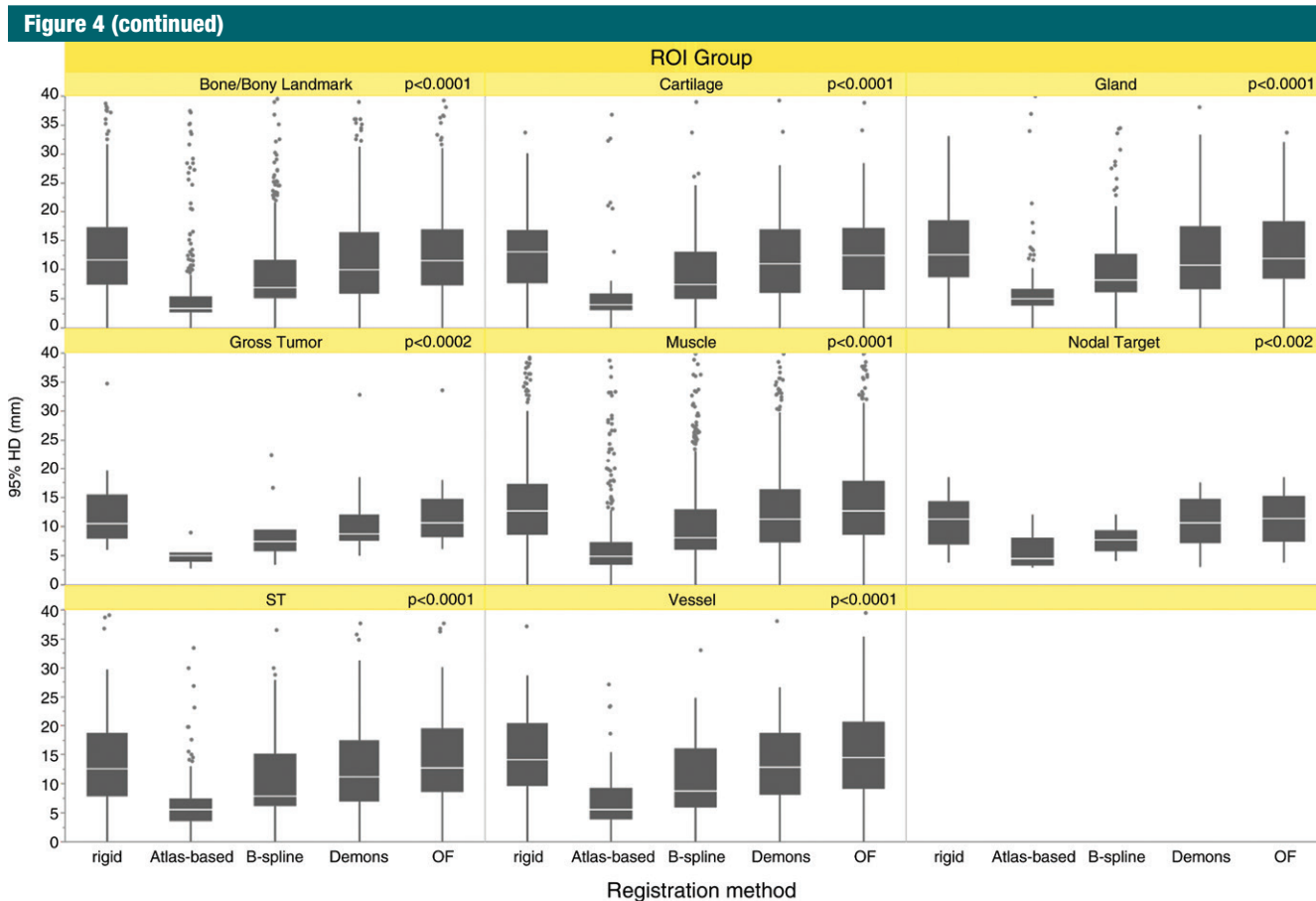
Figure 4 (continued)

b.

Figure 4 *(continued)*, **(b)** Box plots of 95% Haussdorff distance analysis for each registration method are shown according to anatomic ROI category. *OF* = optical flow, *ST* = soft tissues, *95% HD* = 95% Hausdorff distance.

simulation CT and diagnostic CT. STR notably affected some algorithms disproportionately, which was most evident with B-spline image fusions.

Several previous studies have served to validate the use of DIR algorithms in image-guided radiation therapy for head and neck cancer. Approaches used to quantify the performance of distinct DIR in these studies include landmark identification (12), ROI-based comparison (3,10,13,34,35), and computational phantom deformation (8,9,36,37); each of these methods has specific caveats and limitations of application. In the examined setting of diagnostic CT and simulation CT coregistration, the application of an evenly and densely distributed matrix of anatomic landmark points

is intuitively understandable and, with sufficient point placement, exceptionally spatially accurate and statistically robust as a validation method (24). Point placement is, nonetheless, comparatively resource intensive, requiring accurate manual identification of hundreds or thousands of points (24). Point placement in the head and neck is technically complicated secondary to substantial variation in patient position, image acquisition parameter differences (especially STR, which is a substantial limitation of voxelwise point identification), and tubular internal anatomy of soft-tissue ROIs in the neck, which enhances the difficulty of manually placing reproducible matched points in paired diagnostic CT and simulation CT image sets. For this effort,

small-scale landmark point placement (approximately 100 points) was used as an intermediary step in our quality assurance chain for optimization of noncommercial DIR settings; however, large-scale landmark point registration efforts, while underway, are yet to be completed by our group, owing to the limitations listed.

As a part of this effort, we sought to develop a "head and neck QA ROI library," as a robust set of labeled anatomic ROIs, to properly represent the unique characteristics of the head and neck anatomy, which contains multiple structures of different shapes, sizes, and intensity gradients in Hounsfield units. ROI-based assessments, as used in the present series (ie, carefully and rigorously reviewed

Table 3

Effect of Variability in Simulation CT Section Thickness Reconstruction on the Performance of Registration Algorithms

Registration Method	STR < 3 mm*	STR ≥ 3 mm*	P Value
Rigid			
95% Hausdorff distance (mm)	15.7 ± 12.4	14.4 ± 11.5	.25
DSC	0.24 ± 0.22	0.27 ± 0.22	.006†
Atlas based			
95% Hausdorff distance (mm)	7.2 ± 10.6	7.7 ± 11.2	<.0001†‡
DSC	0.64 ± 0.15	0.65 ± 0.14	.74
B-spline			
95% Hausdorff distance (mm)	10.8 ± 10.1	10.5 ± 11.2	.02†
DSC	0.41 ± 0.22	0.47 ± 0.20	<.0001†‡
Demons			
95% Hausdorff distance (mm)	14.3 ± 12.2	13 ± 11.6	.17
DSC	0.33 ± 0.25	0.38 ± 0.24	.0019†‡
Optical flow			
95% Hausdorff distance (mm)	16.1 ± 13.2	14.3 ± 11.1	.14
DSC	0.28 ± 0.22	0.31 ± 0.22	.0034†‡

Note.—Data are means ± standard deviations, unless indicated otherwise.

* $n = 10$ patients.

† Significant $P < .05$ for Wilcoxon rank test.

‡ Significant $P < .005$ after Bonferroni correction ($\alpha = .05/10$ "pairwise comparison of five DIR algorithms").

and curated independently by two radiation oncologists after initial manual segmentation), give insight into changes of shape, volume, and location of a structure. Thus, ROI or ROI category–specific performance differences can be ascertained, a feature that may be overlooked with "anatomically agnostic" DIR QA approaches, such as point registration or image intensity matching. Likewise, the large number of ROIs and ROI categories used in this QA workflow more nearly approximate the range of DIR algorithm performance across anatomic subsites in the head and neck, as opposed to "sample ROI" methods or use of a limited number of reference ROIs, which may underrepresent interregional DIR variation.

Only a few investigators have validated the registration of diagnostic CT and simulation CT images, two of which have addressed the registration of pretherapy diagnostic positron emission tomography/CT images with simulation CT images (15,16), while in two other studies, the registration of simulation CT images with postrecurrence diagnostic CT images

was examined (17,18). The validation methods for DIR assessment in those studies varied from calculating the root mean squared error from an observer's set of marked anatomic landmarks (15) to overlap indexes and center of mass comparison for sample ROIs (16–18). Ireland et al (15) and Hwang et al (16) showed that, in accordance with our results, DIR achieved superior performance to RIR. Additionally, Due et al (18) showed that DIR has higher reproducibility than RIR in repeated registration of center of mass points used to identify the origin of local-regional recurrences mapped to original simulation CT images. These results should call into critical question the utility of RIR as an accurate tool for head and neck registration extracranially, and we advise cautious use of RIR as only a "rough guide" rather than a serially implemented clinical tool for diagnostic CT and simulation CT head and neck workflows. However, although optical flow (12) and demons (3,10,36) algorithms have been studied and have shown utility in adaptive radiation therapy applications, they

failed to evince comparable results in the specific setting of head and neck diagnostic CT and simulation CT image coregistration.

Several limitations must be noted. As a single-institution study that involved the use of retrospective image data, the standard caveats apply. Given this particular anatomic site (head and neck) and the fact that image sets (diagnostic CT and simulation CT) were acquired by using standard institutional operating procedures, excessive generalization regarding DIR algorithm performance in other anatomic sites or differing acquisition settings is potentially specious. Since the development of this QA process and ROI library was exceedingly resource intensive, owing to the time required to manually segment and review a comparatively massive number of anatomic ROI structures, this library of paired diagnostic CT and simulation CT images and ROI structure sets is provided as anonymized DICOM radiation therapy files to any other researchers at *http://figshare.com/authors/Abdallah_Mohamed/551961.* Additionally, the effect of surgical resection or induction chemotherapy on the quality of registration was not investigated in the present study, since we sought to benchmark best performance scenarios and exclude the effect of anatomic distortion caused by huge tumors.

In summary, we developed a QA framework by using a robust, curated, manually segmented anatomic ROI library to quantitatively assess different image registration strategies used for coregistration of head and neck diagnostic CT images with simulation CT images. The presented QA framework proved that DIR algorithms for most of the tested metrics improved registration performance over RIR, yet with notable variability between different algorithms, suggesting that careful validation of DIR before clinical implementation (eg, target delineation) is imperative.

References

1. Cazoulat G, Lesaunier M, Simon A, et al. From image-guided radiotherapy to dose-guided radiotherapy [in French]. Cancer Radiother 2011;15(8):691–698.

2. Grégoire V, Jeraj R, Lee JA, O'Sullivan B. Radiotherapy for head and neck tumours in 2012 and beyond: conformal, tailored, and adaptive? Lancet Oncol 2012;13(7):e292–e300.

3. Hardcastle N, Tomé WA, Cannon DM, et al. A multi-institution evaluation of deformable image registration algorithms for automatic organ delineation in adaptive head and neck radiotherapy. Radiat Oncol 2012;7:90.

4. Ezhil M, Choi B, Starkschall G, Bucci MK, Vedam S, Balter P. Comparison of rigid and adaptive methods of propagating gross tumor volume through respiratory phases of four-dimensional computed tomography image data set. Int J Radiat Oncol Biol Phys 2008;71(1):290–296.

5. Brock KK; Deformable Registration Accuracy Consortium. Results of a multi-institution deformable registration accuracy study (MIDRAS). Int J Radiat Oncol Biol Phys 2010;76(2):583–596.

6. Cui Y, Galvin JM, Straube WL, et al. Multi-system verification of registrations for image-guided radiotherapy in clinical trials. Int J Radiat Oncol Biol Phys 2011;81(1):305–312.

7. Kirby N, Chuang C, Ueda U, Pouliot J. The need for application-based adaptation of deformable image registration. Med Phys 2013;40(1):011702.

8. Nie K, Chuang C, Kirby N, Braunstein S, Pouliot J. Site-specific deformable imaging registration algorithm selection using patient-based simulated deformations. Med Phys 2013;40(4):041911.

9. Pukala J, Meeks SL, Staton RJ, Bova FJ, Mañon RR, Langen KM. A virtual phantom library for the quantification of deformable image registration uncertainties in patients with cancers of the head and neck. Med Phys 2013;40(11):111703.

10. Castadot P, Lee JA, Parraga A, Geets X, Macq B, Grégoire V. Comparison of 12 deformable registration strategies in adaptive radiation therapy for the treatment of head and neck tumors. Radiother Oncol 2008;89(1):1–12.

11. Lee C, Langen KM, Lu W, et al. Evaluation of geometric changes of parotid glands during head and neck cancer radiotherapy using daily MVCT and automatic deformable registration. Radiother Oncol 2008;89(1):81–88.

12. Østergaard Noe K, De Senneville BD, Elstrøm UV, Tanderup K, Sørensen TS. Acceleration and validation of optical flow based deformable registration for image-guided radiotherapy. Acta Oncol 2008;47(7):1286–1293.

13. Huger S, Graff P, Harter V, et al. Evaluation of the Block Matching deformable registration algorithm in the field of head-and-neck adaptive radiotherapy. Phys Med 2014;30(3):301–308.

14. Aird EG, Conway J. CT simulation for radiotherapy treatment planning. Br J Radiol 2002;75(900):937–949.

15. Ireland RH, Dyker KE, Barber DC, et al. Nonrigid image registration for head and neck cancer radiotherapy treatment planning with PET/CT. Int J Radiat Oncol Biol Phys 2007;68(3):952–957.

16. Hwang AB, Bacharach SL, Yom SS, et al. Can positron emission tomography (PET) or PET/Computed Tomography (CT) acquired in a nontreatment position be accurately registered to a head-and-neck radiotherapy planning CT? Int J Radiat Oncol Biol Phys 2009;73(2):578–584.

17. Shakam A, Scrimger R, Liu D, et al. Dose-volume analysis of locoregional recurrences in head and neck IMRT, as determined by deformable registration: a prospective multi-institutional trial. Radiother Oncol 2011;99(2):101–107.

18. Due AK, Vogelius IR, Aznar MC, et al. Methods for estimating the site of origin of locoregional recurrence in head and neck squamous cell carcinoma. Strahlenther Onkol 2012;188(8):671–676.

19. Kaanders JH, Fleming TJ, Ang KK, Maor MH, Peters LJ. Devices valuable in head and neck radiotherapy. Int J Radiat Oncol Biol Phys 1992;23(3):639–645.

20. Han X, Hoogeman MS, Levendag PC, et al. Atlas-based auto-segmentation of head and neck CT images. Med Image Comput Comput Assist Interv 2008;11(Pt 2):434–441.

21. Theodorsson-Norheim E. Kruskal-Wallis test: BASIC computer program to perform nonparametric one-way analysis of variance and multiple comparisons on ranks of several independent samples. Comput Methods Programs Biomed 1986;23(1):57–62.

22. Yoo TS, Ackerman MJ, Lorensen WE, et al. Engineering and algorithm design for an image processing Api: a technical report on ITK—the Insight Toolkit. Stud Health Technol Inform 2002;85:586–592.

23. Castillo E, Castillo R, Zhang Y, Guerrero T. Compressible image registration for thoracic computed tomography images. J Med Biol Eng 2009;29(5):222–233.

24. Castillo R, Castillo E, Guerra R, et al. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Phys Med Biol 2009; 54(7):1849–1870.

25. Kalpathy-Cramer J, Fuller CD. Target Contour Testing/Instructional Computer Software (TaCTICS): a novel training and evaluation platform for radiotherapy target delineation. AMIA Annu Symp Proc 2010;2010:361–365.

26. Kalpathy-Cramer J, Awan M, Bedrick S, Rasch CR, Rosenthal DI, Fuller CD. Development of a software for quantitative evaluation radiotherapy target and organ-at-risk segmentation comparison. J Digit Imaging 2014;27(1):108–119.

27. Awan M, Kalpathy-Cramer J, Gunn GB, et al. Prospective assessment of an atlas-based intervention combined with real-time software feedback in contouring lymph node levels and organs-at-risk in the head and neck: quantitative assessment of conformance to expert delineation. Pract Radiat Oncol 2013; 3(3):186–193.

28. Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26(3):297–302.

29. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. IEEE Trans Pattern Anal Mach Intell 1993;15(9):850–863.

30. Crum WR, Camara O, Hill DL. Generalized overlap measures for evaluation and validation in medical image analysis. IEEE Trans Med Imaging 2006;25(11):1451–1461.

31. Steel RG. A multiple comparison rank sum test: treatments versus control. Biometrics 1959;15(4):560–572.

32. Wilcoxon F. Individual comparisons of grouped data by ranking methods. J Econ Entomol 1946;39:269.

33. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 1952;47(260):583–621.

34. Al-Mayah A, Moseley J, Hunter S, et al. Biomechanical-based image registration for head and neck radiation treatment. Phys Med Biol 2010;55(21):6491–6500.

35. Tsuji SY, Hwang A, Weinberg V, Yom SS, Quivey JM, Xia P. Dosimetric evaluation of automatic segmentation for adaptive IMRT for head-and-neck cancer. Int J Radiat Oncol Biol Phys 2010;77(3):707–714.

36. 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(12):2887–2905.

37. 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(1):4066.