Validation of deformable registration in head and neck cancer using analysis of variance

A. Mencarelli, S. van Beek, S. van Kranen, C. Rasch, M. van Herk, and J.-J. Sonke^{a)}

Department of Radiation Oncology, Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital, Plesmanlaan 121, Amsterdam 1066 CX, The Netherlands

(Received 6 August 2012; revised 27 September 2012; accepted for publication 29 September 2012; published 23 October 2012)

Purpose: Deformable image registration (DIR) is often validated based on a distance-to-agreement (DTA) criterion of automatically propagated anatomical landmarks that were manually identified. Due to human observer variability, however, the performance of the registration method is diluted. The purpose of this study was to evaluate an analysis of variance (ANOVA) based validation to account for such observer variation.

Methods: Weekly cone beam CTs (CBCTs) of ten head and neck cancer patients undergoing five weeks of radiotherapy were used. An expert identified 23 anatomical features (landmarks) on the planning CT. The landmarks were automatically propagated to the CBCT using multiregion-of-interest (mROI) registration. Additionally, two human observers independently localized these landmarks on the CBCTs. Subsequently, ANOVA was used to compute the variance of each observer on the pairwise distance (PWD).

Results: ANOVA based analysis demonstrated that a classical DTA approach underestimated the precision for the mROI due to human observer variation by about 25%. The systematic error (accuracy) of mROI ranged from 0.13 to 0.17 mm; the variability (1 SD) (precision) ranged from 1.3 to 1.5 mm demonstrating that its performance is dominated by the precision.

Conclusions: The PWD-ANOVA method accounts for human observer variation allowing a better estimation of the of DIR errors. © 2012 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4760990>]

Key words: deformable registration validation, analysis of variance, observer variation, head and neck cancer

I. INTRODUCTION

Currently, there is a large interest to use deformable image registration (DIR) methods in radiation therapy for dose accumulation^{1,2} and adaptive replanning strategies.^{3,4} Since such DIR methods are designed to support clinical decisions, the quantification of accuracy and precision become a fundamental part of the process of the DIR development and its clinical implementation.

A validation method applied by many authors is distance-to-agreement (DTA) that calculates the distances between reference landmarks and the corresponding landmarks in a registered image.^{5–12} However, the quality of landmark based validation methods is hampered by the ability with which the landmarks can be identified. Consequently, the quantified accuracy and precision of the DIR method represent the combined variation of the landmark identification and the intrinsic accuracy of the registration. To take the observer variation into account, we applied the pairwise differences (PWD) method,¹³ based on the analysis-of-variance (ANOVA).

The purpose of this work was to evaluate the PWD-ANOVA based validation method for deformable image registration to account for observer variation.

II. MATERIALS AND METHODS

In this retrospective study, ten head and neck (H&N) cancer patients without prior tumor stage or location selection were included who had routinely undergone radical RT with cone beam CT (CBCT) guidance. A total of 78 weekly CBCT scans were used.

To select the patients the image quality of planning CT (pCT) scans was assessed: patients with considerable artifacts in the pCT, e.g., motion artifacts due to swallowing or due to slice thickness changes, were not eligible in this study as we wanted to focus on true anatomical changes. Patients were immobilized with a five point thermoplastic fixation mask with shoulder fixation (Civco Medical Solutions, Kolona) and the head was tilted backwards by means of a patient specific neck rest. Eight of the ten pCT scans were reconstructed with a voxel size of $0.8 \times 0.8 \times 3 \text{ mm}^3$ and two with a voxel size $0.98 \times 0.98 \times 1 \text{ mm}^3$. The CBCT (Elekta Synergy 4.2, Elekta Oncology Systems Ltd, Crawley, UK, augmented with inhouse developed software) scans were acquired with an energy of 120 kV and an isocenter dose of about 1 cGy and reconstructed with a voxel size of $1 \times 1 \times 1 \text{ mm}^3$.

FIG. 1. The identified landmarks, represented by black dots, on planning CT inside the head and neck region between the palatine bone and the jugular notch. Five landmarks on the opposite side are not visible. The list of landmarks is shown in Table I.

II.A. Validation

We identified 23 anatomical features (landmarks) distributed in a region from the palatine bone to the jugular notch (Fig. 1). The landmarks were selected to represent different tissues: the air-soft-tissue interface, the bone-soft-tissue interface, and inhomogeneities of the soft tissues. Eight of 23 landmarks were on bony anatomy and 15 on soft tissue structures (Table I).

One expert observer identified the landmarks on the pCT of each patient and, afterwards, the same observer localized the landmarks on the CBCT scans. A second observer independently identified the anatomical landmarks on the CBCT scans. Subsequently, the landmarks on the pCT were automatically propagated to the CBCT scan using a DIR. To quantify the variability of the human observers (HO) and the algorithm, we applied the ANOVA, as showed below.

II.A.1. Human observers

The HOs used an inhouse software tool to place the landmarks on the CBCT. Although the observers did not repeat the same measurement on the same CBCT, recollection bias may occur if observers sequentially handled CBCT scans of the same patient. Therefore, we designed the following pro-

TABLE I. List of the landmarks: in the first column the identification number (IN) of the landmarks, in accord with the localization order described in the protocol; in the second one the anatomical correspondence.

IN	Anatomical location
1, 2	Left, right dens inferior articular process
3	Ventral middle body C3
4	Dorsal middle spinosus process C7
5, 6	Left, right mandible angle/body
7, 8	First left, right rib angle
9	Uvula top
10	Epiglottis
11, 12	Left, right external jugular vein
13	Dorsal cricoid cartilage
14, 15	Left, right middle neck muscle bifurcation
16, 17	Left, right carotid calcification
18	Philtrum (upper lip)
19, 20	Left, right auricula and head contact point
21	Jugular notch
22, 23	Skin neck left, right

cedure for the landmark identification: First, we created a hard copy of the sagittal, transversal, and coronal planes of the pCT for each landmark position. The hard copy was used by the observers as a reference to identify the corresponding landmarks on the CBCT. Second, the tool duplicated the landmarks from the pCT onto the CBCT dataset adding a random displacement of 5 mm (1 SD) in each direction. In that way, the human observer was always forced to move the landmarks in three directions. Third, the CBCT scans were ordered per fraction: the first set contained only the first fraction of each patient; the second set only the second fraction and so forth. The observers had to complete a set following an ordered patient sequence. The patient order was the same for each set. In this way, the observers did not handle two consecutive CBCT scans of the same patient.

II.A.2. Multiple region-of-interest registration

CBCT scans were registered to pCT scans using multiple region-of-interest registration¹⁴ (mROI), Fig. 2. mROI registration is a method based on local rigid-body registration of multiple ROI surrounding relevant bony anatomical structures. Nine 3D rectangular shaped ROI were placed around bony structures which were identified in the pCT: cervical vertebrae C1, C3, C5, C7, lower mandible, hyoid, larynx, occipital bone, and manubrium. Subsequently, the bony anatomy in each ROI was locally rigidly registered to the CBCT using the chamfer-matching algorithm and the simplex optimizer.¹⁵ Finally, the local registration results were interpolated applying thin-plate splines.¹⁶ Six control points for each ROI were used: the control points were located in the center of each side of the rectangular shaped ROI. The positions of each six control point set were recalculated after the local rigid registration and these drove the thin-plate splines deformation.

FIG. 2. Multiregion-of-interest used for our registration method illustrated on the sagittal plane of CBCT; the bony anatomy in each ROI is rigidly registered and interpolated using thin-plate splines.

II.A.3. Analysis of variance

The observer performance is characterized by its accuracy, defined as the closeness of the results to true value, and its precision, the dispersion of the results with respect to true value. Modeling the observation error of an observer as sum of systematic deviation and random variation, the observer accuracy is defined by the systematic deviation of the observation error and the precision by the random variation.

In order to account for observer variation, we applied the PWD-ANOVA-based method¹³ that calculates the random variation of an individual observer, by examining the variations in the differences between the measurements of the ob-

servers. The three observers are indicated by O_1 for the first human observer, O_2 for the second, and O_3 for the registration method. Each observer $O_{i,j} = (O_{x,i,j}, O_{y,i,j}, O_{z,i,j})$, measured the true i th landmark displacement $\mu_i = (\mu_{x,i}, \mu_{y,i}, \mu_{z,i})$ with an observer error, Fig. 3(a). The measurements were rearranged, Fig. 3(b), in the following way:

$$O_{i,j} = \mu_i + m_j + (\delta_i)_j, \quad \text{with } 1 \leq i \leq N \text{ and } 1 \leq j \leq 3, \quad (1)$$

where $O_{i,j}$ indicates the position of the i th landmark of the j th observer and N the total number of landmarks; μ_i is the measure of true but unknown displacement of the landmark; the sum of the terms $m_j = (m_{x,j}, m_{y,j}, m_{z,j})$, the observer specific systematic error (accuracy) from the true landmark displacement, and $(\delta_i)_j = ((\delta_{x,i})_j, (\delta_{y,i})_j, (\delta_{z,i})_j)$, the random displacement with zero mean represents the observer variation (precision). Since the true landmark displacement is unknown, we measure the accuracy of the observers O_3 relative to the mean of the observer O_1 and O_2 . This is an arbitrary choice of reference that serves as an estimate for the unknown true landmark displacement. We could choose another reference, e.g., the mean of the positions of the three observers or one of the two HO without affecting the random variation analysis.¹³ Computing the pairwise differences between two observations, the unknown true i th landmark displacement μ_i is cancelled out of the equations, Fig. 3(c),

$$\begin{aligned} O_{i,2} - O_{i,1} &= 2m_m + (\delta_i)_2 - (\delta_i)_1, \\ O_{i,3} - O_{i,1} &= m_3 - m_m + (\delta_i)_3 - (\delta_i)_1, \\ O_{i,3} - O_{i,2} &= m_3 - m_m + (\delta_i)_3 - (\delta_i)_2, \end{aligned} \quad (2)$$

where the right terms are the difference of the random displacement of two observers and the term $m_m = (m_{x,m}, m_{y,m}, m_{z,m})$ is the specific systematic error of the observers O_1 and O_2 relative to the estimate of the true landmark displacement. The variance of this difference is given as the sum of the variance of the random variation of each observer, i.e., their precision. This is therefore a system of three linear equations with three unknowns and the solutions of the system give the precision of the three observers, i.e.,

$$\begin{aligned} \sigma_{\delta O_1}^2 &= (\sigma_{O_2 O_1}^2 + \sigma_{O_3 O_1}^2 - \sigma_{O_3 O_2}^2)/2, \\ \sigma_{\delta O_2}^2 &= (\sigma_{O_3 O_2}^2 + \sigma_{O_2 O_1}^2 - \sigma_{O_3 O_1}^2)/2, \\ \sigma_{\delta O_3}^2 &= (\sigma_{O_3 O_1}^2 + \sigma_{O_3 O_2}^2 - \sigma_{O_2 O_1}^2)/2, \end{aligned} \quad (3)$$

FIG. 3. Model for the three observers, for simplicity the accuracy term m is not shown: (a) coordinates of the true displacement, μ_i , and the observer estimated positions of the i th landmark with respect to data set reference system; (b) rearrangement of the same observations; (c) pairwise differences between $O_{i,1}$, $O_{i,2}$, and $O_{i,3}$.

where $\sigma_{O_2O_1}^2$, $\sigma_{O_3O_1}^2$, and $\sigma_{O_3O_2}^2$ are the variances of the differences and $\sigma_{\delta O_1}^2$, $\sigma_{\delta O_2}^2$, and $\sigma_{\delta O_3}^2$ are the variances of the three observers, respectively.

II.A.4. Outliers

The PWD-ANOVA method is sensitive to the presence of outliers.¹³ In our method, the outliers were identified with the values of the distribution of each pairwise difference between two observers that exceeded a threshold of four times the standard deviation (SD). Prior to estimating the variances, the outliers were removed from the measurement and counted. Subsequently, the outliers were analyzed in two ways: (1) per “landmark” to compute the number of outliers for a specific landmark per patient; (2) per outlier “coincidence” in order to separate outliers induced by human observers and the registration algorithm. In case an outlier was found in two observer pairs, the outlier was attributed to the common observer of the two pairs. In case an outlier was present in all three pairs the attribution was not possible. The remaining noncoincident outliers were attributed to the human observers if generated by the human-human observer pair and otherwise to the algorithm.

II.B. DTA comparison

We compared the PWD-ANOVA with the DTA method using the mean of the observations of the two HOs. The outliers computed in the PWD-ANOVA analysis were also removed from the set of DTA measurements for a consistent analysis.

III. RESULTS

III.A. DTA comparison

The observer precision for all-landmarks is presented, together with the standard deviation of the DTA analyses, in Table II. The human observer precision ranged from 0.5 to 1.5 mm and was consistent between the two observers. The accuracy of the mROI, common to PWD-ANOVA and DTA, are given in Table III. All the values were significantly different from zero ($p < 0.001$), however, submillimeter.

These results demonstrate that the random variation is considerably larger than the systematic deviation and that the per-

TABLE III. The accuracy (mm) of multiregion-of-interest registration (mROI) per axis (LR = left-right, CC = cranial-caudal, AP = anterior-posterior) and vector length (3D), for the whole set of landmarks. The underlined values are significantly different from zero.

Observer	All landmarks			
	LR	CC	AP	3D
mROI	<u>0.13</u>	<u>0.17</u>	<u>0.15</u>	<u>0.26</u>

formance of the DIR can thus be described by its precision. We compared ANOVA with the DTA in terms of precision and found that DTA underestimate the precision of the mROI by about 23%, 35%, and 20% in the left-right, cranial-caudal, and anterior-posterior direction, respectively.

III.B. Outliers

In the all-landmarks analysis mROI, generated 55 outliers, representing 3.1% of total numbers of landmarks. 26% of the outliers were on the uvula (landmark 9). The coincidence analysis revealed the human-human pair generated 44% of the total outliers.

To assess the performance of the validation method, we applied Monte Carlo simulations.¹³ With a dataset of 1739 cases, representing the total number of valid observation, randomly sampled from a normal distribution, 10 000 repeated measurements were simulated to compute the standard error of PWD-ANOVA method. For the all-landmarks analysis, we found a standard error of 2.7%.

IV. DISCUSSION

To the best of our knowledge, this is the first study applying PWD-ANOVA for deformable image registration validation. It was shown that human observer variation had a considerable impact on a classical landmark based DTA analysis. Using a PWD-ANOVA based analysis method, the registration precision and human observer variation could be separated.

The choice of the DIR method was arbitrary to the purpose of this study because the scope was to propose a validation method and not to validate a specific DIR algorithm.

IV.A. Landmark approach

Besides DTA of landmarks various other quantitative validation methods have been proposed such as consistency testing,¹⁷ mathematical and similarity metrics,^{6,18} and DTA of 2D/3D contours.^{12,18} For this study, we applied landmarks because they quantify local performance of the registrations in 3D.

In literature, there is no agreement about the minimal number of landmarks that have to be applied and few studies were dedicated to the H&N region. Noe⁸ in the H&N study for validation of CBCT to CT registration posed six landmarks at easy identifiable positions on the bony anatomy. Nithiananthan¹⁰ used eight anatomical points, seven on bony

TABLE II. Observer precision (1 SD, mm) of human observers and the multiregion-of-interest registration (mROI) from ANOVA; standard deviation for mROI from DTA per axis (LR = left-right, CC = cranial-caudal, AP = anterior-posterior) and mean vector length (3D), for the whole set of landmarks.

Statistical method	Observer	All landmarks			
		LR	CC	AP	3D
ANOVA	First observer	1.0	1.5	1.2	2.1
	Second observer	0.5	0.8	0.8	1.2
	mROI	1.3	1.3	1.5	2.4
DTA	mROI	1.6	1.8	1.8	3.0

anatomy and one on soft tissues, either epiglottis or uvula, for CBCT to CT registration in the H&N. Malsch⁵ manually selected circa 78 landmarks in two CTs of two H&N patients.

IV.B. Human observers

The analysis of outliers revealed that HO variations gave a substantial contribution to the outliers. These outliers were due to either misinterpretations or differences in judgment between the two HOs.

The human-human pairwise observers presented 24 outliers: eight on the bony landmarks and 16 on the soft-tissue. All the bony outliers were located on the ribs and were likely caused by breathing artifacts. In the soft-tissues, there were no landmarks that presented a concentration of outliers. The cause of soft-tissue human outliers may be attributed to low soft-tissue contrast, presence of motion artifacts due to swallowing and cupping artifacts.

The HO variability was, however, low and consistent between the two observers as well as with findings of other authors. Kashani⁷ used 48 plastic markers, with a diameter of 2.5 mm, imbedded in a lung phantom; one observer repeated identification of a subset of 10 markers the same markers made by a single observer and he reported an average precision ranging from 0.4 to 0.8 mm in RL, AP, and SI direction.

IV.C. Outlier detection

In this study, the outliers were detected using a 4 SD threshold on the Euclidian distance between two observers. An alternative approach is to apply a multivariate outlier detection method based on the Mahalanobis distance¹⁹ that takes the covariance between the observers into account. Using the F-distribution for the Mahalanobis distance, with a critical value at the 5% significance level, very similar results were obtained with circa 10% less outliers and a maximal difference in estimated precision < 0.1 mm.

IV.D. Observer variation and ground truth

To assess the observer variation, a straightforward method is to repeat the observations of the same cases and to compute the SD of the findings.⁷ However, observer recollection may introduce a bias. To avoid the ground truth problem, different approaches are applied, regardless of the chosen validation method: synthetic data,²⁰ simulated deformations,²¹ digital phantoms,²² and physical phantoms.^{7,9} Synthetic data, simulated deformation, and digital phantoms define the ground truth but it is difficult to identify and model realistic sources of variability. With physical phantoms the ground truth is easily obtained from the measures of the phantom before and after mechanical deformations. However, the performances are evaluated in specific conditions and they do not include anatomical complexity and variation. Sarrut⁶ used two or more observers to estimate the ground truth. In the study of Sarrut,⁶ a pseudo ground truth, defined as the mean of the measurements of four observers, was used. The PWD-ANOVA method extends this idea separating the random and

systematic errors of individual observers, with the advantage of requiring less HOs. Each user identifies the landmarks only once, consequently the intra-HO bias is avoided.

Although using the same reference as in PWD-ANOVA is a logical choice, by computing the DTA in relation of the mean of two observers the impact of the observer variation is reduced: with the first HO, for example, the DTA underestimation of the precision ranged from 26% to 49%.

We demonstrated (Table III) that the observer accuracy were submillimeter and that the observer uncertainties are dominated by precision. Additionally, we have shown that the human variation is a relevant part of the DIR total errors in the H&N.

IV.E. Limitations of the study

In the presented method, the systematic deviations of the observer O_3 are measured relative to the mean of the two human observers. This implicates that, first, the ground truth problem is fully avoided only for the random errors; second, the model of systematic deviation does not assure a correct computation of the DIR algorithm accuracy limiting the PWD-ANOVA method to measure only the precision. However, the systematic deviation of the DIR computed with respect to the average of the two HOs were submillimeter, i.e., not clinically relevant. These results together with the DIR precision may be used to fully characterize the DIR performance.

The PWD-ANOVA method is sensitive to the presence of outliers and these were removed before applying the method. Although these outliers are not directly included in the performance quantification, they are analyzed to identify DIR algorithm local failures and they may point out the causes of the algorithm limitations in different clinical applications.

To perform the validation a large number of measurements are necessary to assure a high precision and accuracy of the method itself.

V. CONCLUSIONS

We proposed a validation method based on PWD-ANOVA to quantify the accuracy and precision of deformable image registration that accounts for observer variation. The method separates the random and systematic errors of individual observers. As an example, we evaluated deformable registration of cone-beam CT scan of H&N cancer patients and found that a classical distance-to-agreement approach underestimated the performance of our registration method due to the contribution of human observer variation by about 25%.

^{a)}Author to whom correspondence should be addressed. Electronic mail: janjakob@nki.nl

¹B. Schaly, J. A. Kempe, G. S. Bauman, J. J. Battista, and J. Van Dyk, "Tracking the dose distribution in radiation therapy by accounting for variable anatomy," *Phys. Med. Biol.* **49**, 791–805 (2004).

²J. Orban de Xivry *et al.*, "Tumour delineation and cumulative dose computation in radiotherapy based on deformable registration of respiratory correlated CT images of lung cancer patients," *Radiother. Oncol.* **85**, 232–238 (2007).

- ³W. Lu, G. H. Olivera, Q. Chen, M. L. Chen, and K. J. Ruchala, "Automatic re-contouring in 4D radiotherapy," *Phys. Med. Biol.* **51**, 1077–1099 (2006).
- ⁴M. Chao, Y. Xie, and L. Xing, "Auto-propagation of contours for adaptive prostate radiation therapy," *Phys. Med. Biol.* **53**, 4533–4542 (2008).
- ⁵U. Malsch, C. Thieke, P. E. Huber, and R. Bendl, "An enhanced block matching algorithm for fast elastic registration in adaptive radiotherapy," *Phys. Med. Biol.* **51**, 4789–4806 (2006).
- ⁶D. Sarrut, V. Boldea, S. Miguët, and C. Ginestet, "Simulation of four-dimensional CT images from deformable registration between inhale and exhale breath-hold CT scans," *Med. Phys.* **33**, 605–617 (2006).
- ⁷R. Kashani, M. Hub, M. L. Kessler, and J. M. Balter, "A physical phantom for assessment of accuracy of deformable alignment algorithms," *Med. Phys.* **34**, 2785–2788 (2007).
- ⁸K. Ø. Noe, B. D. de Senneville, U. V. Elstrøm, K. Tanderup, and T. S. Sørensen, "Acceleration and validation of optical flow based deformable registration for image-guided radiotherapy," *Acta Oncol.* **47**, 1286–1293 (2008).
- ⁹M. Serban, E. Heath, G. Stroian, D. L. Collins, and J. Seuntjens, "A deformable phantom for 4D radiotherapy verification: Design and image registration evaluation," *Med. Phys.* **35**, 1094–2002 (2008).
- ¹⁰S. Nithianathan, K. K. Brock, M. J. Daly, H. Chan, J. C. Irish, and J. H. Siewerdsen, "Demons deformable registration for CBCT-guided procedures in the head and neck: Convergence and accuracy," *Med. Phys.* **36**, 4755–4764 (2009).
- ¹¹R. Castillo *et al.*, "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets," *Phys. Med. Biol.* **54**, 1849–1870 (2009).
- ¹²K. K. Brock *et al.*, "Results of a multi-institution deformable registration accuracy study (MIDRAS)," *Int. J. Radiat. Oncol. Biol. Phys.* **76**, 583–596 (2010).
- ¹³K. G. Gilhuijs, A. Touw, M. van Herk, and R. E. Vijlbrief, "Optimization of automatic portal image analysis," *Med. Phys.* **22**, 1089–1099 (1995).
- ¹⁴S. van Kranen, S. van Beek, C. Rasch, M. van Herk, and J.-J. Sonke, "Setup uncertainties of anatomical sub-regions in head-and-neck cancer patients after offline CBCT guidance," *Int. J. Radiat. Oncol. Biol. Phys.* **73**, 1566–1573 (2009).
- ¹⁵M. van Herk and H. M. Kooy, "Automatic three-dimensional correlation of CT-CT, CTMRI, and CT-SPECT using chamfer matching," *Med. Phys.* **21**, 1163–1178 (1994).
- ¹⁶F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 567–585 (1989).
- ¹⁷G. E. Christensen and H. J. Johnson, "Consistent image registration," *IEEE Trans. Med. Imaging* **20**, 568–582 (2001).
- ¹⁸P. Castadot, J. A. Lee, A. Parraga, X. Geets, B. Macq, and V. Grègoire, "Comparison of 12 deformable registration strategies in adaptive radiation therapy for the treatment of head and neck tumors," *Radiother. Oncol.* **89**, 1–12 (2008).
- ¹⁹K. I. Penny, "Appropriate critical values when testing for a single multivariate outlier by using the Mahalanobis distance," *J. R. Stat. Soc., Ser. C, Appl. Stat.* **45**(1), 73–81 (1996).
- ²⁰S. Klein, M. Staring, and J. P. W. Pluim, "Comparison of gradient approximation techniques for optimisation of mutual information in nonrigid registration," *Proc. SPIE* **5747**, 192–203 (2005).
- ²¹H. Wang, L. Dong, J. O'Daniel, R. Mohan, A. S. Garden, K. K. Ang, D. A. Kuban, M. Bonnen, J. Y. Chang, and R. Cheung, "Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy," *Phys. Med. Biol.* **50**, 2887–2905 (2005).
- ²²H. Zhong, J. Kim, and I. J. Chetty, "Analysis of deformable image registration accuracy using computational modeling," *Med. Phys.* **37**, 970–979 (2010).