See SunSCAN 3D in action:
Request your demo today

Introducing SunSCAN™ 3D

The Next-Generation Cylindrical Water Scanning System

SunSCAN 3D simplifies beam scanning with SRS-class accuracy and user-centered design.

It enables faster, easier workflows, and hyper-accurate dosimetry for today's busy clinics.

Learn more:
sunnuclear.com

SUN NUCLEAR
A MIRION MEDICAL COMPANY

SunSCAN™ 3D is not available for sale in all markets. CE Mark pending.

Received: 21 March 2022 | Revised: 9 May 2022 | Accepted: 28 June 2022

DOI: 10.1002/mp.15851

RESEARCH ARTICLE

MEDICAL PHYSICS

Skull phantom-based methodology to validate MRI co-registration accuracy for Gamma Knife radiosurgery

Ryan T. Oglesby[1,2] | **Wilfred W. Lam**[1] | **Mark Ruschin**[3,4] | **Lori Holden**[3,4] | **Arman Sarfehnia**[1,3,4,6] | **Collins Yeboah**[3,4] | **Arjun Sahgal**[3,4] | **Hany Soliman**[3,4] | **Jay Detsky**[3,4] | **Chia-Lin Tseng**[3,4] | **Sten Myrehaug**[3,4] | **Zain Husain**[3,4] | **Angus Z. Lau**[1,2] | **Greg J. Stanisz**[1,2,5,6] | **Brige P. Chugh**[3,4,6]

[1]Physical Sciences, Sunnybrook Research Institute, Toronto, Ontario, Canada

[2]Medical Biophysics, University of Toronto, Toronto, Ontario, Canada

[3]Odette Cancer Centre, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

[4]Department of Radiation Oncology, University of Toronto, Toronto, Ontario, Canada

[5]Neurosurgery and Paediatric Neurosurgery, Medical University of Lublin, Lublin, Poland

[6]Department of Physics, Ryerson University, Toronto, Ontario, Canada

Correspondence
Ryan Oglesby, Physical Sciences, Sunnybrook Research Institute, 2075 Bayview Ave., Room S6 05, Toronto, ON M4N 3M5, Canada.
Email: ryan.oglesby@mail.utoronto.ca

Funding information
Canadian Institutes of Health Research, Grant/Award Number: #PJT156252

Abstract

Purpose: Target localization, for stereotactic radiosurgery (SRS) treatment with Gamma Knife, has become increasingly reliant on the co-registration between the planning MRI and the stereotactic cone-beam computed tomography (CBCT). Validating image registration between modalities would be particularly beneficial when considering the emergence of novel functional and metabolic MRI pulse sequences for target delineation. This study aimed to develop a phantom-based methodology to quantitatively compare the co-registration accuracy of the standard clinical imaging protocol to a representative MRI sequence that was likely to fail co-registration. The comparative methodology presented in this study may serve as a useful tool to evaluate the clinical translatability of novel MRI sequences.

Methods: A realistic human skull phantom with fiducial marker columns was designed and manufactured to fit into a typical MRI head coil and the Gamma Knife patient positioning system. A series of "optimized" 3D MRI sequences—T_1-weighted Dixon, T_1-weighted fast field echo (FFE), and T_2-weighted fluid-attenuated inversion recovery (FLAIR)—were acquired and co-registered to the CBCT. The same sequences were "compromised" by reconstructing without geometric distortion correction and re-collecting with lower signal-to-noise-ratio (SNR) to simulate a novel MRI sequence with poor co-registration accuracy. Image similarity metrics—structural similarity (SSIM) index, mean squared error (MSE), and peak SNR (PSNR)—were used to quantitatively compare the co-registration of the optimized and compromised MR images.

Results: The ground truth fiducial positions were compared to positions measured from each optimized image volume revealing a maximum median geometric uncertainty of 0.39 mm (LR), 0.92 mm (AP), and 0.13 mm (SI) between the CT and CBCT, 0.60 mm (LR), 0.36 mm (AP), and 0.07 mm (SI) between the CT and T_1-weighted Dixon, 0.42 mm (LR), 0.23 mm (AP), and 0.08 mm (SI) between the CT and T_1-weighted FFE, and 0.45 mm (LR), 0.19 mm (AP), and 1.04 mm (SI) between the CT and T_2-weighted FLAIR. Qualitatively, pairs of optimized and compromised image slices were compared using a fusion image where separable colors were used to differentiate between images. Quantitatively, MSE was the most predictive and SSIM the second most predictive metric for evaluating co-registration similarity. A clinically relevant threshold of MSE, SSIM, and/or PSNR may be defined beyond which point an MRI sequence should be rejected for target delineation based on its dissimilarity

to an optimized sequence co-registration. All dissimilarity thresholds calculated using correlation coefficients with in-plane geometric uncertainty would need to be defined on a sequence-by-sequence basis and validated with patient data.
Conclusion: This study utilized a realistic skull phantom and image similarity metrics to develop a methodology capable of quantitatively assessing whether a modern research-based MRI sequence can be co-registered to the Gamma Knife CBCT with equal or less than equal accuracy when compared to a clinically accepted protocol.

KEYWORDS
cone beam computed tomography, co-registration, gamma knife, magnetic resonance imaging, radiation therapy, skull phantom

1 | INTRODUCTION

Stereotactic radiosurgery (SRS) is a radiation therapy technique, which utilizes precisely focused photon beams to deliver a highly concentrated dose to a target area, often in a single treatment session, with high spatial accuracy and steep dose gradients to minimize injury to the surrounding healthy structures. Gamma Knife radiosurgery is a cranial SRS technique, often hypo-fractionated that requires sub-millimeter accuracy in image guidance to ensure safe radiation delivery to patients.[1] For brain metastases, the target volume is typically defined using 3D T_1-weighted magnetic resonance imaging (MRI) with gadolinium contrast enhancement; edema is defined using 3D T_2-weighted MRI; and a geometrically accurate reference image is obtained using computed tomography (CT).[2] Gamma Knife offers cone-beam CT (CBCT) on-board imaging as stereotactic reference.[3] Target localization for SRS has become increasingly reliant on the co-registration between the nonstereotactic planning MRI and stereotactic CBCT.[4] Currently, co-registration accuracy is qualitatively verified in the treatment planning software by overlaying the planning MRI on the CBCT and manually ensuring coincidence of skull-based landmarks. This process is sensitive to MRI parameters, particularly those that relate to geometric distortion, such as B_0 strength, receiver bandwidth, and gradient nonlinearity correction.[5] Patient-specific parameters including motion artifacts and/or susceptibility artifacts (e.g., permanent retainers) may additionally contribute to an imperfect co-registration.

With the continuously evolving optimization of MRI pulse sequences for target delineation comes the need to validate image registration between modalities, particularly when considering the use of emerging functional and metabolic imaging techniques,[6] such as magnetization transfer (MT) sequences used to distinguish radiation necrosis from tumour progression.[7] Evaluating the co-registration accuracy of modern research-based MRI sequences in relation to the standard clinical MRI sequences used for Gamma Knife is critical to ensure that modern sequences are being co-registered to the CBCT with at least equivalent accuracy when compared to clinical sequences.

Building on previous work[8] and relevant Gamma Knife co-registration literature,[9–12] this study aimed to develop a phantom-based methodology to quantitatively compare the co-registration accuracy of a compromised MRI sequence with unknown signal and geometric variability to a standard clinical MRI sequence optimized for target delineation. Commercially available MRI distortion phantoms are limited in their ability to validate modern research-based MRI sequences for the Gamma Knife workflow since they do not accurately model the geometry, image contrast, and distortion artifacts observed in the skull of a patient. To overcome these limitations, a realistic skull phantom was required, which mimicked some of the challenges in co-registration between MRI and CBCT.

In this study, a realistic skull phantom was designed and built to accurately model the geometry and distortion artifacts typically observed in bone. The skull phantom was imaged using a variety of MRI sequences, which were each separately co-registered to a CBCT volume. Three clinical standard 3D MRI sequences—T_1-weighted Dixon, T_1-weighted fast field echo (FFE), and T_2-weighted fluid-attenuated inversion recovery (FLAIR)—were defined as "optimized" sequences for image co-registration. For each optimized sequence, two corresponding "compromised" sequences were acquired, one reconstructed without geometric distortion correction and the other with low signal-to-noise ratio (SNR). The compromised sequences represent two extreme cases of modern research-based MRI sequences, which are likely to fail when co-registered to the CBCT. The optimized and compromised sequences were compared slice-by-slice after co-registration using various quantitative image similarity metrics to evaluate co-registration similarity. The methods presented in this study provide a quantitative measure for comparing the co-registration of a modern research-based MRI sequence to a similar clinical standard sequence, such as the MRI protocol described above for Gamma Knife.

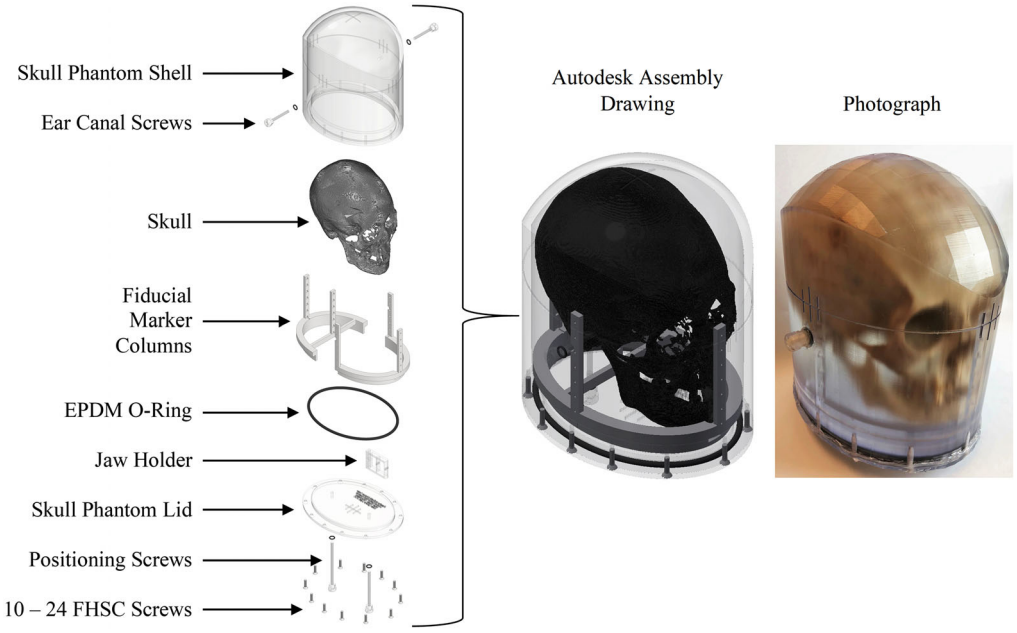

FIGURE 1 Skull phantom assembly. Left: Annotated exploded view. Right: Assembly view and photograph. The skull and mandible were suspended in the phantom using two ear canal screws, two positioning screws, and a jaw holder. Five fiducial marker columns—located on each of the four sides of the skull and one protruding through the foramen magnum—were used to aid in the assessment of geometric uncertainty and co-registration similarity. The skull phantom shell was filled with deionized water, and the lid was sealed using 10–24 flat head socket cap (FHSC) nylon screws, an ethylene propylene diene monomer (EPDM) O-ring, and silicone sealant

This comparative methodology may serve as a valuable tool to evaluate the clinical translatability of novel MRI sequences for SRS.

2 | METHODS

2.1 | Skull phantom design

A human skull was purchased (Osta International, White Rock, BC, Canada), which had been cleaned and bleached but not further chemically treated in order to maintain the native MR properties of bone. A skull surface model was created from CT data and used as a template to design the geometry of the skull phantom (Figure 1) in Autodesk Inventor Pro 2020 (Autodesk, San Rafael, CA, USA). The size and shape of the skull phantom was minimized to allow for easy insertion into a typical MRI head coil as well as the patient positioning system on the Gamma Knife Icon (Elekta, Stockholm, Sweden). This meant that the skull phantom had to be contained within a geometry of approximately 17 cm left-right and 22 cm anterior-posterior. The superior-inferior dimension was a less relevant design constraint because the inferior end of any head positioning system is open to make room for the neck. A skull phantom shell, fiducial marker columns, jaw holder, and lid were 3D printed (Viper si, 3D Systems, Rock Hill, SC, USA) out of Accura ClearVue plastic. Five fiducial marker columns for landmarking were included in the design to aid in the quantitative assessment of geometric distortion and co-registration, one located at each of the four sides of the skull and one penetrating through the foramen magnum. Seven identical fiducial subunits, each consisting of three cut out rectangular sticks ($2 \times 2 \times 5$ mm^3) aligned along the Cartesian axes, were equally distributed along each column forming 35 fiducial markers visible on both MRI and CT. The skull was suspended in the shell using four nylon screws (one in each ear canal and two on the inferior side of the skull) to ensure positioning reproducibility and minimize contact with the walls. Additionally, a two-piece jaw holder was designed, and 3D printed to hold the mandible in position. The skull phantom was filled with deionized water, and the lid was sealed using 10–24 flat head socket cap (FHSC) nylon screws, an ethylene propylene diene monomer (EPDM) O-ring, and silicone sealant.

2.2 | Image acquisition

A 3D spiral CT volume (Brilliance Big Bore, Philips Medical Systems, Best, Netherlands) was acquired with 120 kVp, 415 mA, FOV = $350 \times 350 \times 250$ mm^3, 1 mm slice thickness, and 512×512 in-plane resolution. A high SNR CBCT volume (CTDI = 6.3 mGy) was acquired using the Gamma Knife Icon (Elekta, Stockholm,

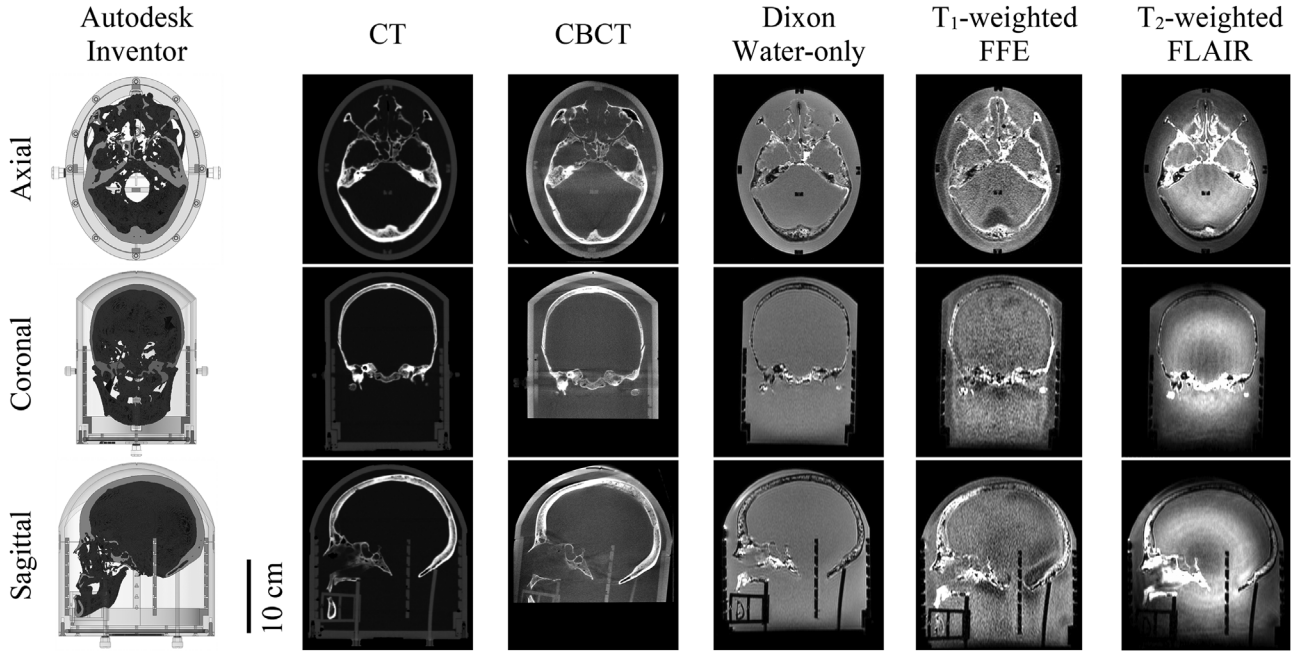

FIGURE 2 Through plane (axial, coronal, and sagittal) image slices of the skull phantom from the Autodesk Inventor design, computed tomography (CT), cone-beam computed tomography (CBCT), and Dixon water-only, T_1-weighted fast field echo (FFE), and T_2-weighted fluid-attenuated inversion recovery (FLAIR) MRI sequences

Sweden) scanner with 90 kVp, FOV = 224 × 224 × 224 mm^3, 0.5-mm slice thickness, and 448 × 448 in-plane resolution. MR images were acquired at 3T (Achieva, Philips Medical Systems, Best, Netherlands) using a Q-body transmit coil and a SENSE-Head-8 receive coil. Three different 3D axial MRI volumes were reconstructed: a T_1-weighted Dixon-derived water-only volume; a T_1-weighted FFE volume; and a T_2-weighted FLAIR volume. The Dixon volume was acquired using a fast field echo sequence (FFE; TR = 6.629 ms; TE_1/TE_2 = 1.93/3.4 ms; FA = 15°; FOV = 220 × 220 × 240 mm^3; acquired matrix = 220 × 220 × 240; reconstructed matrix = 560 × 560 × 240; pixel bandwidth = 1061 Hz; averages = 2). The T_1-weighted volume was also acquired using a fast field echo sequence (FFE; TR = 4.414 ms; TE = 1.448 ms; FA = 30°; FOV = 220 × 220 × 240 mm^3; acquired matrix = 220 × 220 × 240; reconstructed matrix = 240 × 240 × 240; pixel bandwidth = 803 Hz; averages = 1). Lastly, the FLAIR volume was acquired using a turbo inversion recovery sequence (TIR; TR = 4800 ms; TE = 262.2 ms; echo train length = 182; TI = 1650 ms; FA = 90°; FOV = 220 × 220 × 240 mm^3; acquired matrix = 200 × 200 × 240; reconstructed matrix = 288 × 288 × 120; pixel bandwidth = 1411 Hz; averages = 2). An axial, coronal, and sagittal slice of each modality and MRI sequence is given alongside the skull phantom Autodesk design in Figure 2.

First, in addition to those MRI volumes reconstructed using the standard clinical sequence parameters (denoted as the "optimized" volumes), each volume was reconstructed without the default geometric distortion correction. Second, for comparison, each sequence was acquired again using the Q-body receive coil rather than the SENSE-Head-8 receive coil to demonstrate a case with lower SNR. These additional data are denoted as the "compromised" volumes further in the manuscript. The compromised MRI volumes with low SNR or no geometric distortion correction represent two extreme cases of modern research-based MRI sequences (e.g., echo planar imaging [EPI]-based sequences sensitive to field inhomogeneities), which are likely to fail when co-registered to the CBCT.

2.3 | Data analysis

The first step in the data analysis was to assess the geometric uncertainty between the ground truth positions of the skull phantom fiducials and the positions measured from each MRI volume. The ground truth positions were defined using the high-resolution CT with voxel volumes of 0.68 × 0.68 × 1 mm^3, which are sufficiently precise to serve as the ground truth for MR images collected at 1 mm in-plane spatial resolution. Material deflection due to loading with the skull and water was not observed at the resolution of the images acquired. Fiducial positions in each volume were measured manually using Leksell GammaPlan treatment planning software (LGP; version 11.1.0, Elekta, Stockholm, Sweden), and distances

between each fiducial were calculated and compared to the ground truth to quantify the geometric uncertainty of the optimized MRI volumes. Manual selection of the fiducial positions has an in-plane uncertainty limited by the spatial resolution and a through-plane uncertainty of ±50% slice thickness due to partial volume effects at the interface between water and the fiducials markers.

Next, each MRI sequence was imported into LGP and co-registered to the CBCT using a mutual information algorithm with subvoxel accuracy.[13] The LGP mutual information algorithm finds the best rigid transformation (six degrees of freedom – three rotations and three translations) between anatomically corresponding positions in two images by optimizing the narrowness of an intensity histogram produced using the normalized mutual information measure.[14] The resulting transformation matrices were manually extracted from the LGP file for each MRI sequence. A transformation matrix for the *i*th MRI volume, $\mathbf{A}_i$, was applied to each image in the volume as follows:

$$\mathrm{MR}\ I_i^{'} = \mathbf{A_i}\ \mathrm{MRI}_i, \tag{1}$$

where $\mathrm{MRI}_i^{'}$ is the *i*th MRI volume in the new coordinate system and MRI_i, in the original coordinate system. Rather than directly compare the transformation matrices, we considered it more valuable to transform each MRI and compare the similarity between pairs of optimized and compromised image slices to provide a regional assessment of those locations around the skull that may most negatively impact co-registration. Evaluation of the transformation matrices alone would not provide the spatially specific co-registration errors relevant for treatment planning. The image processing algorithm for each pair of MR images is illustrated in Figure 3.

The final step of image analysis was to assess the similarity between pairs of optimized and compromised MRI sequence slices following co-registration to CBCT stereotactic space. The similarity between pairs of images was assessed using mean squared error (MSE) and peak signal-to-noise ratio (PSNR),[15] as well as the structural similarity (SSIM) index.[16]

If x_i represents the voxel intensity of an MR image to be measured and y_i the voxel intensity of a reference MR image, then the MSE is defined as:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{N}[x_i - y_i]^2}{N}, \tag{2}$$

where N is the number of voxels. The MSE provides a measure of the mean squared difference between the reference image (optimized MRI) and the image to be measured (compromised MRI). In general, a smaller MSE implies fewer differences between images. PSNR is defined as:

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{x_{i,\mathrm{MAX}}^2}{\mathrm{MSE}}\right), \tag{3}$$

where $x_{i,\mathrm{MAX}}$ is the maximum voxel value of the input image. The PSNR is additionally useful in this context because the MR images being compared have different dynamic range. A simple SNR calculation was additionally included for reference and is defined as:

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\overline{y_i^2}}{\mathrm{MSE}}\right), \tag{4}$$

where $\overline{y_i^2}$ is the mean squared reference image. Lastly, the SSIM index is defined as:

$$\mathrm{SSIM}\ (x, y) = [l\,(x, y)]^{\alpha}\ [c\,(x, y)]^{\beta}[s\,(x, y)]^{\gamma} \tag{5}$$

where the SSIM is a weighted combination of luminance (l), contrast (c), and structure (s):

$$l\ (x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{6}$$

$$c\ (x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{7}$$

$$s\ (x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}. \tag{8}$$

If $\alpha = \beta = \gamma = 1$ (equivalent weighting of luminance, contrast, and structure) and $C_3 = C_2/2$, then the SSIM index simplifies to:

$$\mathrm{SSIM}\ (x, y) = \frac{(2\mu_x\mu_y + C_1)\,(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)\,(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{9}$$

where μ_x and μ_y are the mean of x_i and y_i, respectively; ${\sigma_x}^2$ and ${\sigma_y}^2$ are the variance of x_i and y_i, respectively; $C_1 = (k_1 x_{i,\mathrm{MAX}})^2$ and $C_2 = (k_2 x_{i,\mathrm{MAX}})^2$ are constants used to avoid instability when ${\mu_x}^2 + {\mu_y}^2$ is close to zero; k_1 and k_2 are small constants $\ll 1$; and σ_{xy} is the covariance of x_i and y_i defined as:

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y). \tag{10}$$

The SSIM index served as a useful tool to quantify the visibility of co-registration errors between an optimized and compromised image using the known properties of human visual perception: luminance, contrast, and

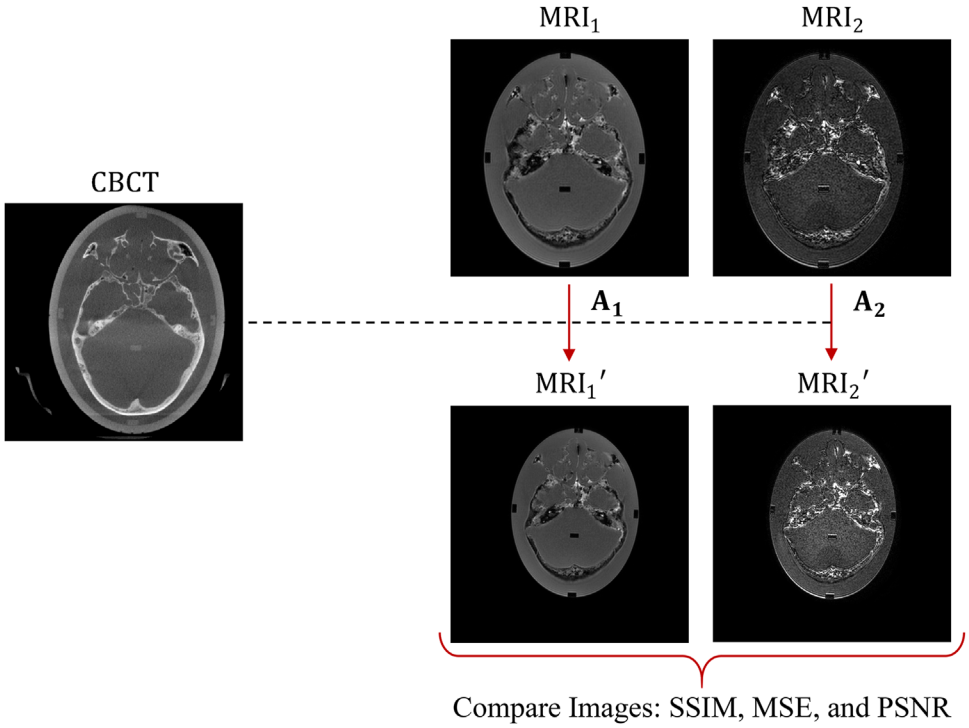

FIGURE 3 Image processing algorithm for each pair of optimized and compromised MR images. Both MRI_1 and MRI_2 are transformed to cone-beam computed tomography (CBCT) stereotactic space and the resultant similarity between images calculated using the structural similarity (SSIM) index, mean squared error (MSE), and peak signal-to-noise ratio (PSNR)

structure. Pearson's correlation coefficients were calculated between SSIM, MSE, and the median geometric uncertainty between the ground truth fiducial positions and the optimized MR image volumes.

3 | RESULTS

The outcome of image comparison between optimized and compromised MRI sequences following co-registration relied on the geometric uncertainty of the optimized images. The ground truth fiducial positions—as defined by the high-resolution CT—were compared to positions measured from each optimized volume: CBCT, Dixon water-only, T_1-weighted FFE, and T_2-weighted FLAIR (Figure 4). The fiducials, consisting of five subunits per axial plane, define eight line segments in the left-right (LR) and anterior-posterior (AP) directions for comparison: AB, AC, CE, AD, BE, BC, CD, and DE. Superior-inferior (SI) geometric uncertainty was estimated along each of the five fiducial marker columns by defining six additional line segments: 1 – 2, 2 – 3, 3 – 4, 4 – 5, 5 – 6, and 6 – 7. The distance between each fiducial, along each line segment, was calculated, and the results are illustrated as a series box plots.

The results of the fiducial position comparison for the optimized images reveal a maximum median geometric uncertainty of 0.39 mm (LR), 0.92 mm (AP), and 0.13 mm (SI) between the CT and CBCT, 0.60 mm (LR), 0.36 mm (AP), and 0.07 mm (SI) between the CT and Dixon water-only, 0.42 mm (LR), 0.23 mm (AP), and 0.08 mm (SI) between the CT and T_1-weighted FFE, and 0.45 mm (LR), 0.19 mm (AP), and 1.04 mm (SI) between the CT and T_2-weighted FLAIR.

Following the quantification of geometric uncertainty in the optimized images, pairs of optimized and compromised images were compared slice-by-slice using the SSIM index, MSE, and PSNR. The differences between pairs of images were qualitatively illustrated by the use of a fusion image where the two images are overlaid using separable colors. In this case, the optimized images are shown in red, and the compromised images are shown in cyan. A fusion montage is illustrated in Figure 5, showing 15 of 500 total slices in CBCT stereotactic space of the Dixon water-only images with and without geometric distortion correction. From these data, it is evident that image distortion and misalignment is most clear at large distances away from magnetic isocenter (defined as slice 0 in CBCT stereotactic space). These misalignments are most dramatically observed at slice location -175 in Figure 5.

In addition, Figure 6 qualitatively illustrates the difference between all optimized and compromised MRI sequences using a fusion image for a selection of axial slices through the skull phantom. Examination of

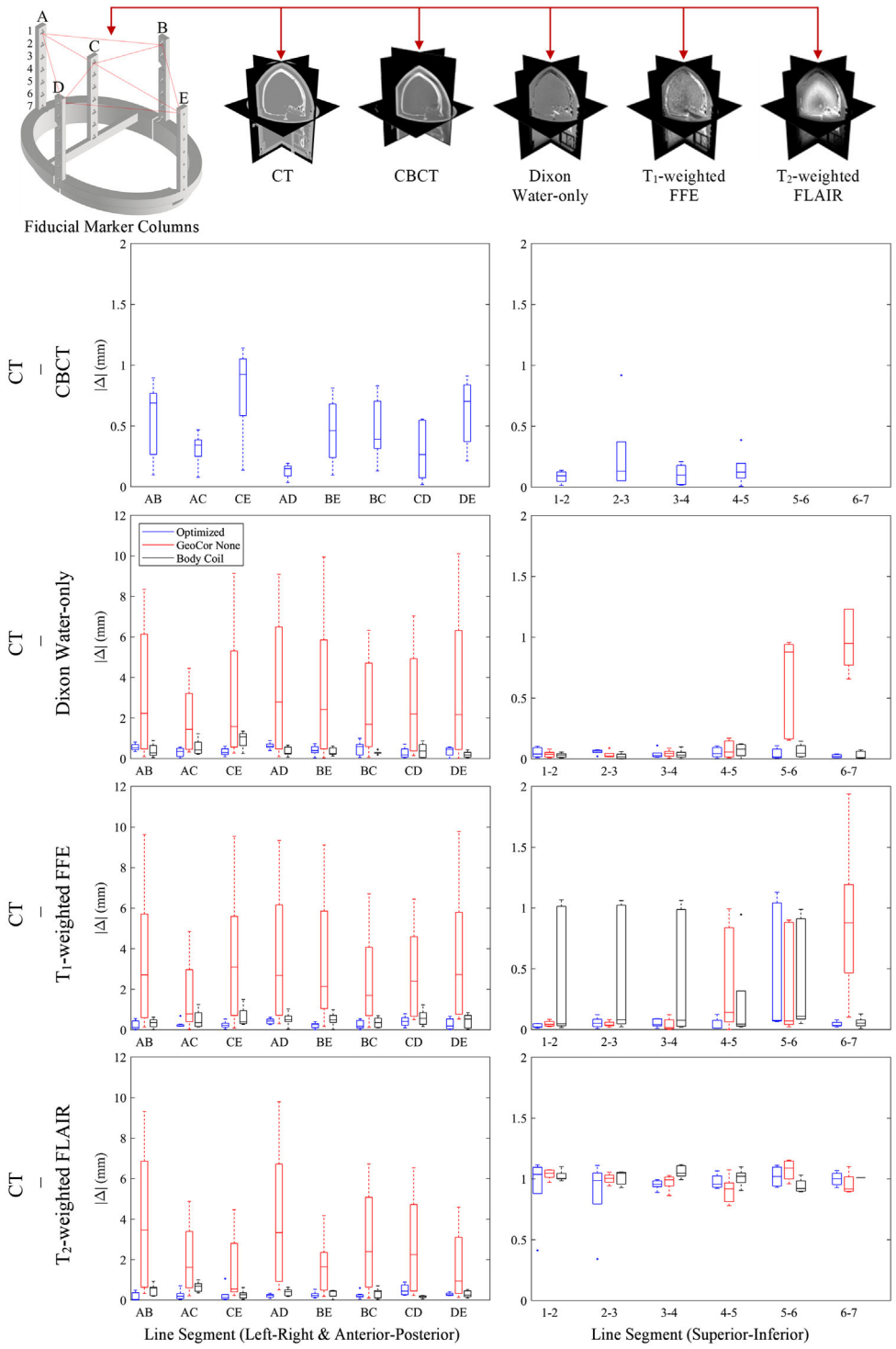

FIGURE 4 Absolute geometric uncertainty |Δ| between the ground truth fiducial positions as defined by the computed tomography (CT) and the optimized images. Top: Fiducial marker column line segment definitions and optimized image volumes: CT, cone-beam computed tomography (CBCT), Dixon water-only, T_1-weighted fast field echo (FFE), and T_2-weighted fluid-attenuated inversion recovery (FLAIR). Bottom: Box plot diagrams demonstrating the difference in line segment length for an optimized image compared to the CT. Left-right and anterior-posterior line segments contains an $N = 7$ sample size, while superior-inferior data contains an $N = 5$ sample size. Note that line segment 6–7 was included for completeness but should not be taken into consideration for geometric uncertainty since the inferior end of the skull phantom was not captured by the CBCT and the signal-to-noise ratio (SNR) of the T_2-weighted FLAIR was too low to accurately determine the location of fiducial marker 7

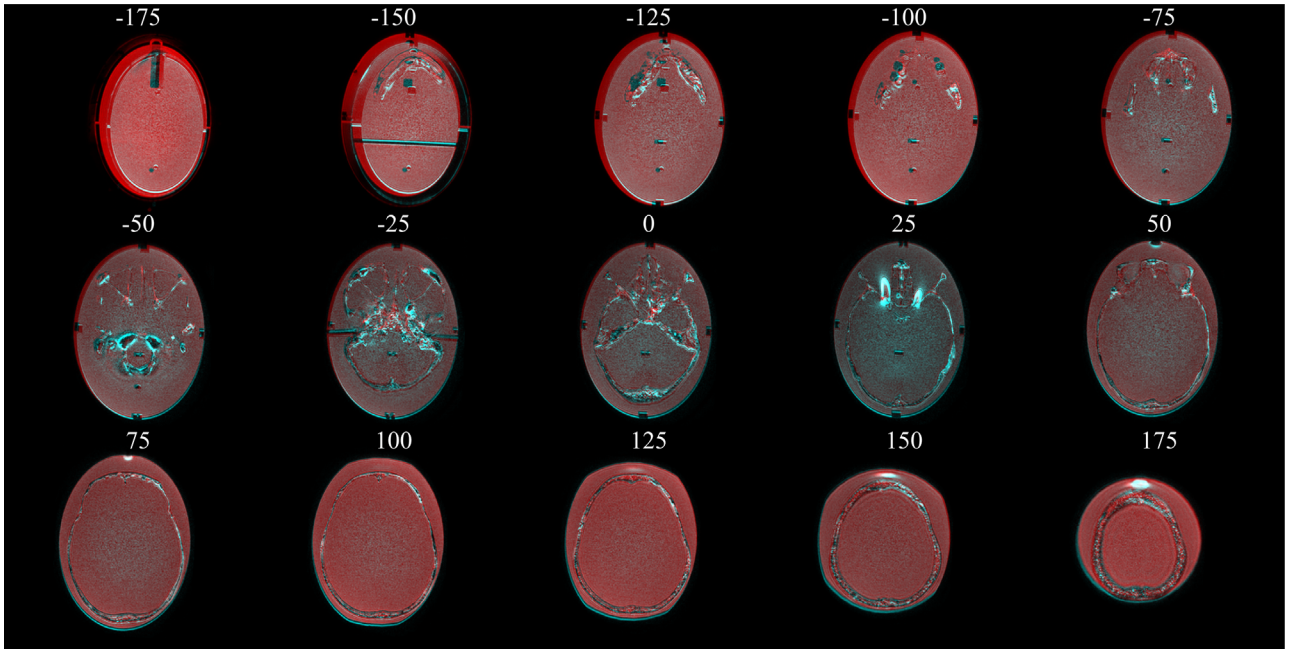

FIGURE 5 Fusion montage illustrating the difference between co-registered Dixon water-only images with geometric distortion correction (red) and without geometric distortion correction (cyan) for a selection of 15 of 500 axial slices through the skull phantom. Magnetic isocenter is co-incident with slice location 0 in cone-beam computed tomography (CBCT) stereotactic space

Figure 6 demonstrates that of the two selected compromised volumes, the volumes reconstructed without geometric distortion correction have a much greater effect on co-registration accuracy than the low SNR volumes collected with the body receive coil. From Figure 6, it is difficult to qualitatively conclude which of the three MRI sequence co-registrations is most negatively affected by the compromization methods. A quantitative image similarity metric is required for this type of assessment.

To demonstrate a more quantitative approach to comparing the optimized and compromised co-registered images, the SSIM index and MSE for each slice are plotted in Figure 7 and the PSNR and SNR, in Figure 8. Given two identical images, the SSIM index would be equal to one and the MSE, to zero. At the edges of the image volumes, where slice location is near ±200, the difference between optimized and compromised images is negligible since there is low signal and high noise in these regions. The SSIM index and MSE become more meaningful and interpretable in the location of the skull phantom with the greatest number of structural details, since it is these details that produce the greatest difference when co-registered. The PSNR, in a reciprocal manner to MSE, is large when the difference between images is low. As PSNR decreases, the difference between images increases. The SNR of the optimized images is included in Figure 8 to aid the interpretation of results. The mean of each similarity metric provided in Table 1 was computed to summarize the difference between pairs of co-registered images as a single global measure. In order to provide a connection between the physical geometric uncertainty defined in Figure 4 and image similarity metrics, Pearson's correlation coefficients were calculated between the median in-plane geometric uncertainty, SSIM index, and MSE in Figure 9. Correlation coefficients for the SSIM index are consistent and range from $R = 0.76–0.81$ (Figure 9a–c), while correlation coefficients for the MSE are inconsistent with a difficult to interpret range from $R = -0.81–0.08$ (Figure 9d–f).

4 | DISCUSSION

This study demonstrated a phantom-based methodology, which may be used to quantitatively assess the similarity between two MRI volumes that have both been co-registered to a CBCT volume. Any co-registered MRI sequence, which deviates beyond an acceptable threshold of MSE, SSIM, and/or PSNR, may be rejected for SRS target delineation on the basis of poor co-registration to the CBCT. All predetermined thresholds would need to be defined on a sequence-by-sequence basis and validated using patient data. The realistic skull phantom developed for this study allows for repeated iterations of MRI sequence optimization prior to defining a clinical dissimilarity threshold. One method of defining an MSE, SSIM, and PSNR threshold would be to first quantify the baseline dissimilarity between

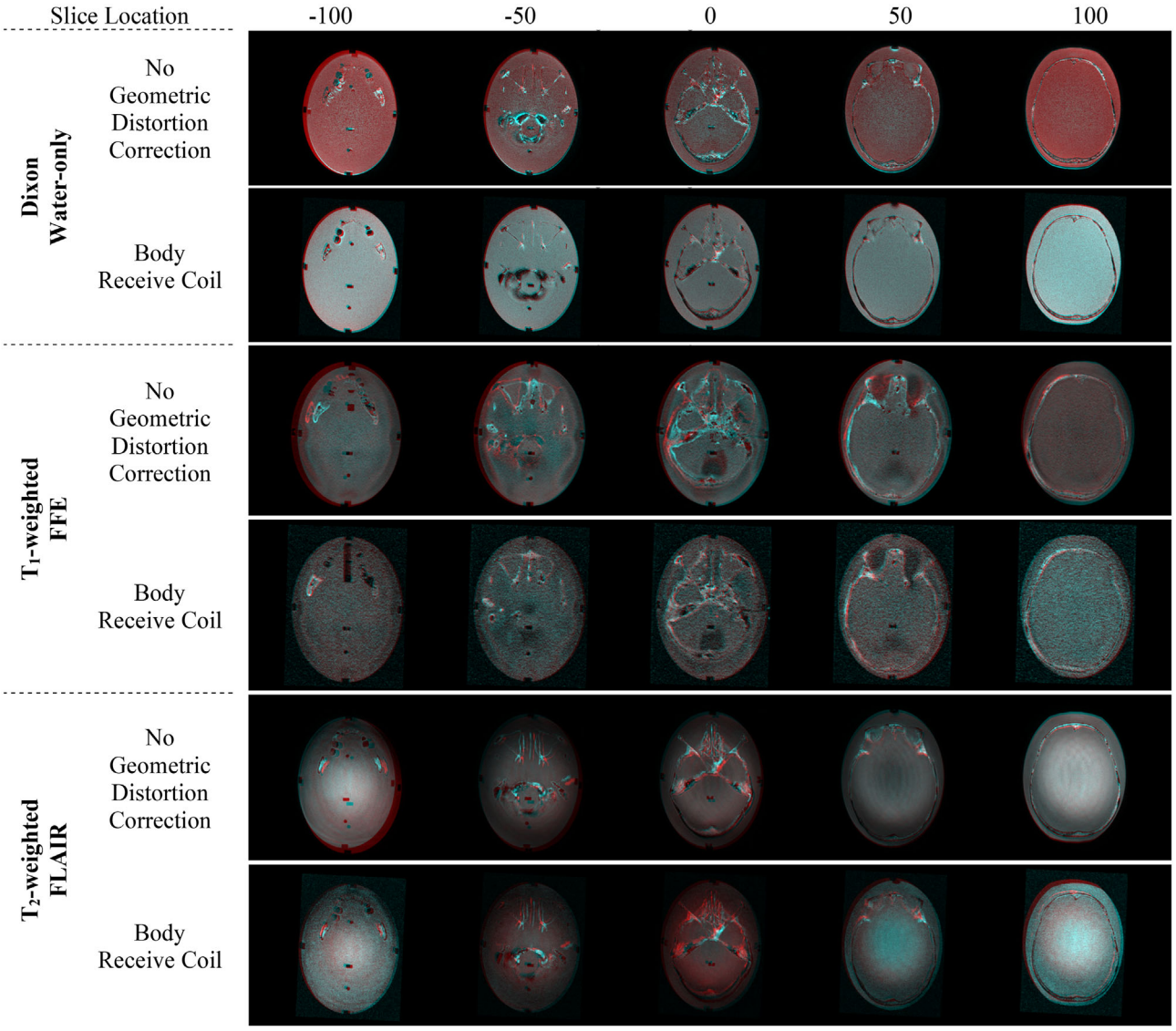

FIGURE 6 Fusion montage illustrating the difference between each optimized (red) and compromised (cyan) co-registered MRI sequence for a selection of 6 of 500 axial slices through the skull phantom. Magnetic isocenter is co-incident with slice location 0 in cone-beam computed tomography (CBCT) stereotactic space

two identical MRI sequences collected sequentially, thereby characterizing the intrinsic dissimilarity caused by MR parameter fluctuation and biological variability. After quantifying the baseline dissimilarity between MRI volumes, one could additionally define an acceptable co-registration-specific dissimilarity based on the correlation with in-plane geometric uncertainty that would adhere to the safety margins of the prescribed dose delivery plan. Defining a clinically relevant dissimilarity threshold will require dose simulations on a variety of co-registered MRI volumes and will be the subject of future work.

Each pair of images compared in Figures 7 and 8 presented an exaggerated case in which an MRI volume co-registration may be rejected on the basis of dissimilarity to an optimized co-registration. The images compromised by reconstructing the data without geometric distortion correction measure a minimum SSIM index value of 0.9993 and a maximum MSE value of 28440. Comparatively, the images compromised by using the body receive coil to reduce SNR measure a minimum SSIM index value of 0.9997 and a maximum MSE value of 25880. By qualitative inspection of the fusion images, we would suggest that the images compromised by reconstructing the data without geometric distortion correction should be rejected, while the images compromised by low SNR could be accepted, based on their similarity to the optimized co-registered MRI sequences. The qualitative suggestion to reject the data reconstructed without geometric distortion correction may be

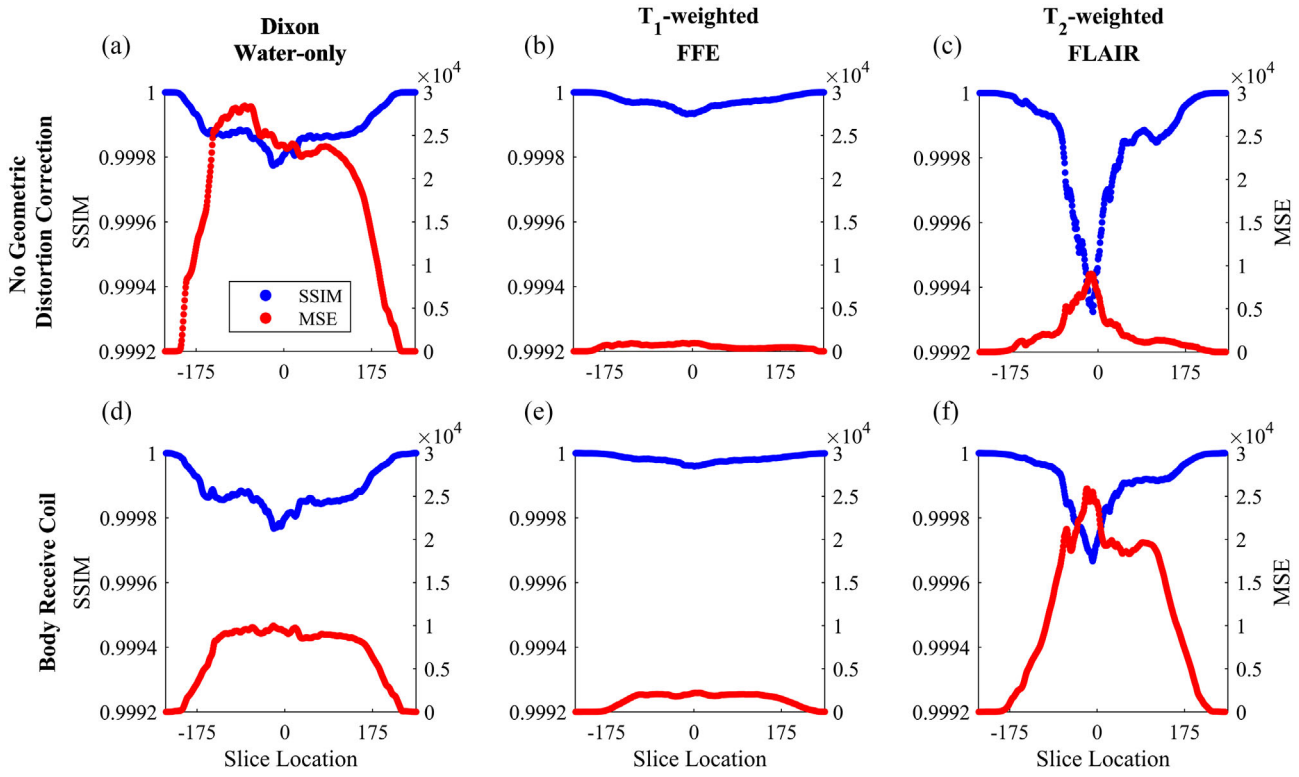

FIGURE 7 Structural similarity (SSIM) index and mean squared error (MSE) as a function of slice location for each pair of optimized and compromised MRI sequence

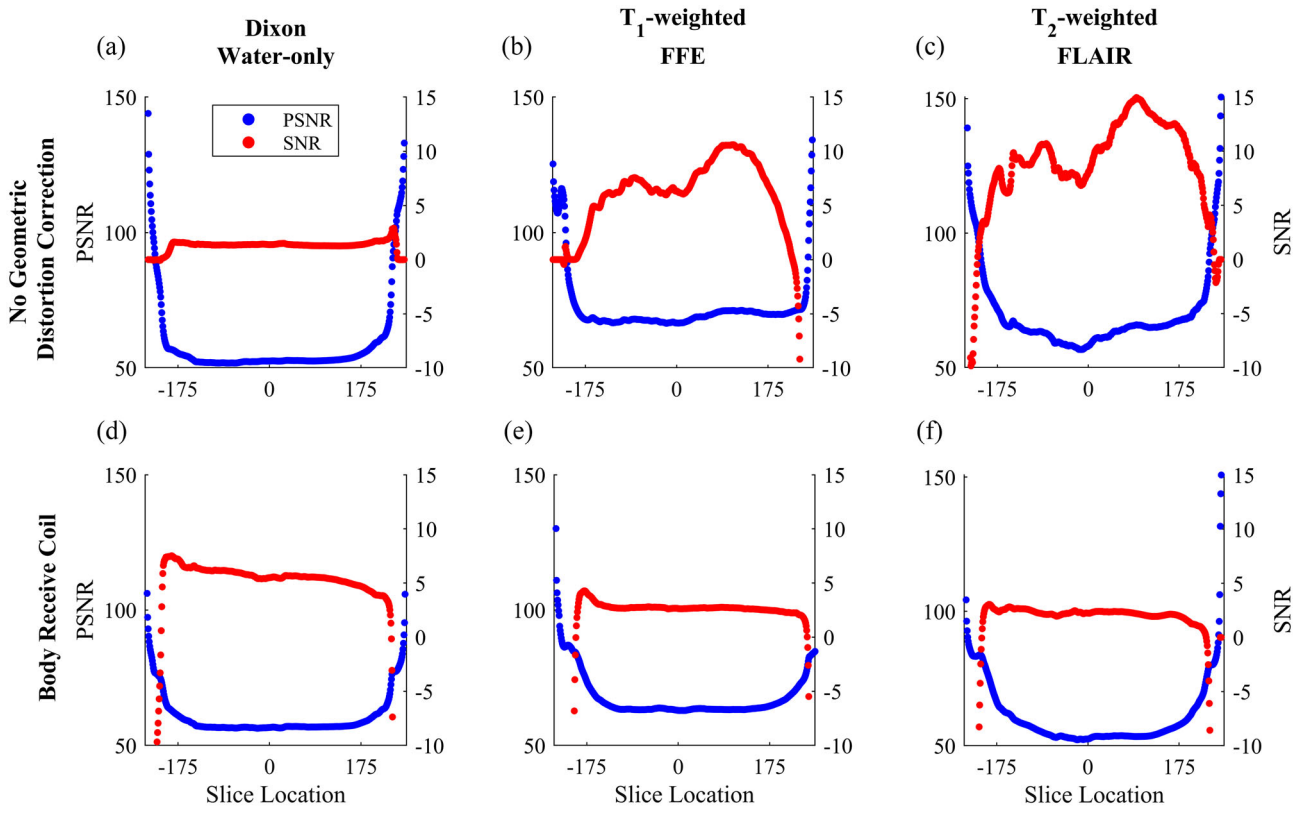

FIGURE 8 Peak signal-to-noise ratio (PSNR) and signal-to-noise ratio (SNR) as a function of slice location for each pair of optimized and compromised MRI sequence

TABLE 1 Mean image similarity metrics between each pair of optimized and compromised co-registered MRI sequences. The voxels were defined by the number of common voxels between the two volumes after co-registration to the cone-beam computed tomography (CBCT)

		Dixon water-only	T_1-weighted FFE	T_2-weighted FLAIR
No geometric distortion correction	SSIM	0.999896	0.999975	0.999860
	MSE	17420	503	2010
	SNR	1.28	4.92	8.64
	PSNR	59.1	72.6	69.9
	Voxels	588 × 493 × 501	507 × 469 × 521	542 × 491 × 495
Body receive coil	SSIM	0.999890	0.999984	0.999923
	MSE	6511	1292	11257
	SNR	2.77	0.55	−0.30
	PSNR	60.8	69.3	61.1
	Voxels	588 × 489 × 501	504 × 470 × 533	566 × 494 × 496

Abbreviations: FFE, fast field echo; FLAIR, fluid-attenuated inversion recovery; MSE, mean squared error; PSNR, peak SNR; SNR, signal-to-noise-ratio; SSIM, structural similarity.

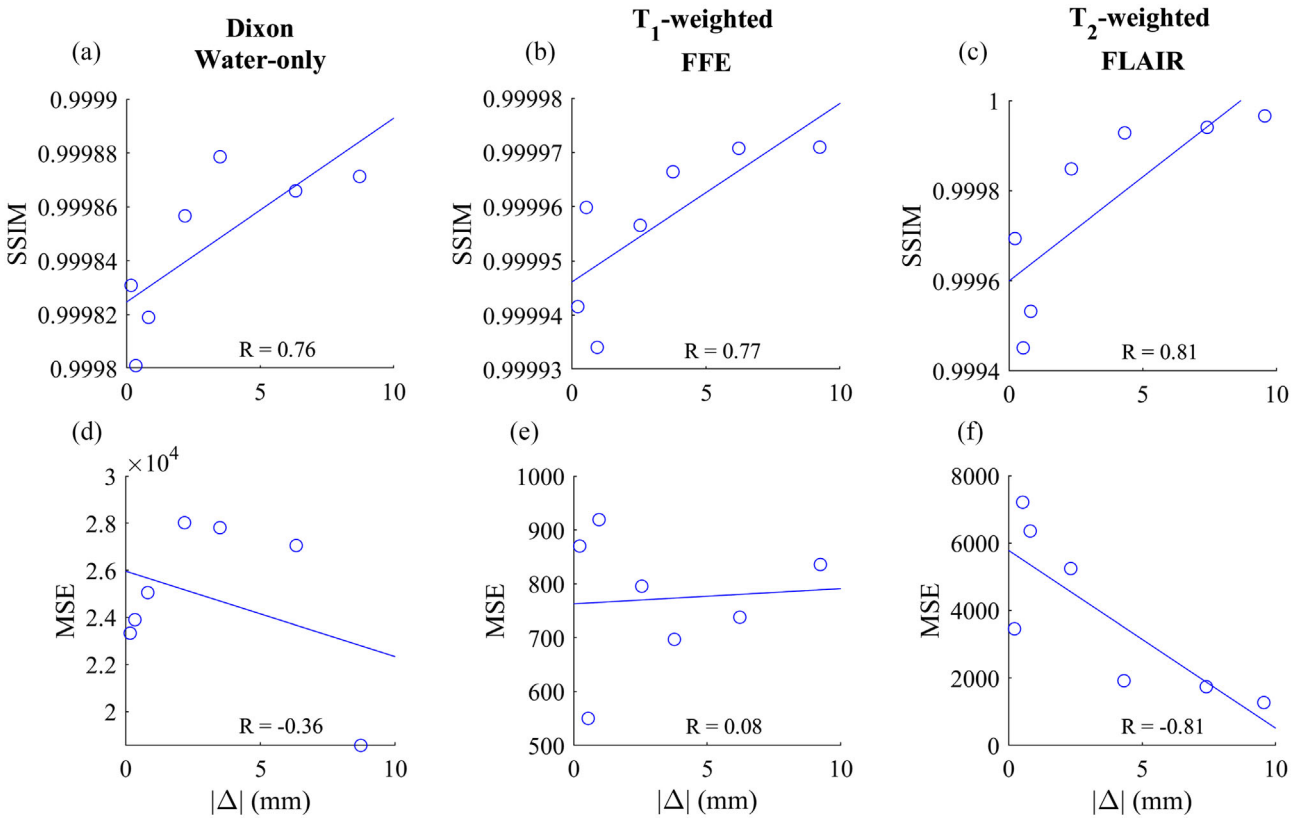

FIGURE 9 Pearson's correlation coefficients comparing structural similarity (SSIM) and mean squared error (MSE) versus median in-plane geometric uncertainty for each MRI sequence compromised without geometric distortion correction

quantitatively supported by the correlation coefficients calculated between SSIM, MSE, and in-plane geometric uncertainty in Figure 9. Correlation coefficients for SSIM are relatively high (R = 0.76–0.81), but for MSE is relatively low (R = −0.81– 0.08) and inconsistent. SSIM and MSE are both sensitive to in-plane geometric uncertainty; however, these metrics are also sensitive to in-plane structural complexity, luminance distortion, and image contrast. Therefore, it is not expected that SSIM and MSE would be perfectly correlated with in-plane geometric uncertainty. Patient data with greater sample sizes should be used to characterize these correlation coefficients more accurately in a clinical setting.

Table 1 presented global image similarity metric means between each pair of optimized and compromised MRI sequences, which may be used as a tool to compare between the co-registration accuracy of compromisation methods. However, for the most accurate assessment of the co-registration differences between the optimized and compromised image volumes, the whole of the similarity metric data must be considered since a mean can hide relevant increases in image dissimilarity. Of each of the image similarity metrics investigated in this study, the MSE appears to be the most predictive of poor co-registration to the CBCT. Of the three clinical standard MRI sequences chosen, the T_1-weighted FFE sequence was the least susceptible to co-registration inaccuracy caused by geometric distortion (mean MSE = 503) and low SNR (mean MSE = 1292); the Dixon water-only, the most susceptible to geometric distortion (mean MSE = 17420); and the T_2-weighted FLAIR sequence, the most susceptible to low SNR (mean MSE = 11257). Note that the methods presented in this study were limited because of their reliance on the accurate co-registration of the optimized clinical MRI sequences to the CBCT. The image similarity metrics were not specific towards the co-registration accuracy of the optimized sequences, but rather indicated if the compromised sequences are co-registered with equal or less than equal accuracy to the optimized sequences.

A critical consideration for future work will be to assess how accurately our compromization methods truly model a novel MRI sequence, and whether there will always exist a clinically optimized MRI sequence for relevant comparison. For example, if the goal were to assess whether a diffusion-weighted sequence could be co-registered to a CBCT with equivalent accuracy when compared to a clinically optimized sequence, two questions arise: (1) what clinically optimized sequence should be chosen for the most accurate co-registration comparison, and (2) will voxel intensity used to calculate SSIM and MSE be sufficiently similar to accurately compare sequences? These two questions must be considered for each novel MRI sequence investigated, and the solutions may vary in complexity depending on the novelty of the sequence being investigated. However, we would suggest for the majority of sequences currently being investigated for target delineation, a clinically optimized sequence exists which can be used for comparison to quantify co-registration accuracy.

The skull phantom utilized in this study provides a more anatomically accurate skull geometry and served its purpose well as a tool for method development. However, it is not necessarily superior to commercially available phantoms at this stage. Commercially available MRI head phantoms come in a variety of styles including: (1) sphere-shaped system phantoms, which aim to standardize T_1 and T_2 relaxation, proton density, and resolution[17–19]; (2) sphere-shaped phantoms, which aim to standardize the blood-oxygen-level-dependent (BOLD) signal,[20] magnetic resonance spectroscopy (MRS),[21] and diffusion-weighted imaging[22,23]; (3) head-shaped end-to-end phantoms, which aim to verify dose delivery[24,25]; and (4) synthetic skull-containing head-shaped phantoms, which aim to mimic realistic anatomy and tissue contrast.[26–29] When compared to commercially available MRI head phantoms, the skull phantom used in this study presented greater similarity in anatomy and native MR properties of bone to that observed in a patient. However, the skull phantom remains limited in its ability to model realistic tissue contrast, air cavities, and pulsatile motion. These limitations can be overcome by further customization of the skull phantom. Realistic tissue contrast can be modeled by combining T_1 and T_2 relaxation contrast agents (e.g., salts of paramagnetic ions) with carbohydrate-containing gels (e.g., agar), rather than the deionized water used in these experiments. Realistic tissue contrast would greatly improve the ability of the skull phantom to access the co-registration accuracy of MRI sequences, which aim to suppress fat signal or quantify relaxivity. The skull phantom could additionally be improved by incorporating air cavities, which mimic the geometry of the sinuses. This could be achieved by inserting a thin walled, hollow (air-filled) part, 3D printed to fit within the geometry of the nasal cavity. The addition of air cavities would improve the ability of the skull phantom to assess co-registration accuracy of MRI sequences sensitive to susceptibility artifacts such as EPI.[30] Similarly, the skull phantom could be modified to evaluate MRI sequences, which aim to reduce metal artifacts such as metal artifact reduction sequence.[31] This could be accomplished by placing a stainless steel or titanium-alloy pedicle screw in a relevant location within the skull phantom. Lastly, and most challenging of all, would be to improve the accuracy of the skull phantom by recreating the pulsatile motion of the basal ganglia and brain stem during systole.[32] It is unclear what the best method to accomplish pulsatile motion would be, but it would most likely require a re-design of the skull phantom to incorporate an MR compatible mechanism to displace contrast agent into and out of the foramen magnum. If pulsatile motion was successfully modeled, then cine MRI sequence co-registration could be evaluated for the purposes of adaptive radiotherapy. However, a simple geometry may be more beneficial for motion evaluation at this stage of development.

Future work will involve repeating the described method of image similarity analysis to assess co-registration of the planning MRI sequences to the CBCT on patient data. The assessment of patient data will require a statistical analysis to account for the greater number of image volumes compared to the single sample volumes presented here. Following the completion of a statistical analysis on patient data and subsequent validation of the proposed method to assess

co-registration accuracy, the next step will be to investigate the most recently developed MRI sequences for target delineation. Noninvasive MRI contrast mechanisms such as diffusion tensor imaging,[33] MT,[34] chemical exchange saturation transfer,[35] magnetic resonance thermometry,[36] and MRS[37] are of particular interest based on recent research in the assessment of radiation therapy effectiveness. If successful, the methodology presented in this study may provide insight into the most valuable MR parameters and/or sequences required for successful co-registration to CBCT, thereby maximizing the accuracy of target localization to achieve the highest quality patient care with Gamma Knife radiosurgery.

5 | CONCLUSIONS

The methods presented in this study utilized image similarity metrics to analyze whether a modern research-based MRI sequence can be co-registered to a Gamma Knife CBCT with equal or less than equal accuracy when compared to a clinically accepted MRI sequence. By extracting the transformation matrices from GammaPlan, it was shown that optimized and compromized MRI sequence co-registration to CBCT can be quantitatively compared using the SSIM index, MSE, and PSNR. A consistent correlation was found between median in-plane geometric uncertainty and SSIM.

ACKNOWLEDGMENTS

The authors thank the Canadian Institutes of Health Research (grant #PJT156252) for funding.

CONFLICT OF INTEREST

Chia-Lin (Eric) Tseng belongs to the Elekta MR-Linac Research Consortium and has received travel accommodations/expenses and honoraria for past educational seminars by Elekta.

DATA AVAILABILITY STATEMENT

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request

REFERENCES

1. Desai R, Rich KM. Therapeutic role of Gamma Knife stereotactic radiosurgery in neuro-oncology. *Mo Med*. 2020;117(1):33-38. http://www.ncbi.nlm.nih.gov/pubmed/32158047
2. Taghizadeh S, Labuda C, Yang CC, et al. Optimizing MRI sequences and images for MRI-based stereotactic radiosurgery treatment planning. *Reports Pract Oncol Radiother*. 2019;24(1):12-19. https://doi.org/10.1016/j.rpor.2018.09.010
3. Sarfehnia A, Ruschin M, Chugh B, et al. Performance characterization of an integrated cone-beam CT system for dedicated gamma radiosurgery. *Med Phys*. 2018;45(9):4179-4190. https://doi.org/10.1002/mp.13073
4. Ruschin M, Sahgal A, Soliman H, et al. Clinical image coregistration variability on a dedicated radiosurgery unit. *Neurosurgery*. 2019;85(1):E101-E108. https://doi.org/10.1093/neuros/nyy334
5. Pappas EP, Alshanqity M, Moutsatsos A, et al. MRI-related geometric distortions in stereotactic radiotherapy treatment planning: evaluation and dosimetric impact. *Technol Cancer Res Treat*. 2017;16(6):1120-1129. https://doi.org/10.1177/1533034617735454
6. Kashani R, Olsen JR. Magnetic resonance imaging for target delineation and daily treatment modification. *Semin Radiat Oncol*. 2018;28(3):178-184. https://doi.org/10.1016/j.semradonc.2018.02.002
7. Mehrabian H, Detsky J, Soliman H, Sahgal A, Stanisz GJ. Advanced magnetic resonance imaging techniques in management of brain metastases. *Front Oncol*. 2019;9(June):440. https://doi.org/10.3389/fonc.2019.00440
8. Soliman AS, Burns L, Owrangi A, et al. A realistic phantom for validating MRI-based synthetic CT images of the human skull. *Med Phys*. 2017;44(9):4687-4694. https://doi.org/10.1002/mp.12428
9. Chung H-T, Kim JH, Kim JW, et al. Assessment of image co-registration accuracy for frameless gamma knife surgery. *PLoS One*. 2018;13(3):e0193809. https://doi.org/10.1371/journal.pone.0193809
10. Nakazawa H, Mori Y, Komori M, et al. Validation of accuracy in image co-registration with computed tomography and magnetic resonance imaging in Gamma Knife radiosurgery. *J Radiat Res*. 2014;55(5):924-933. https://doi.org/10.1093/jrr/rru027
11. Nix M, Wright G, Fallows P, Sykes W, Bownes P. A high-precision, geometric and registration accuracy full-system test method for adaptive SRS, demonstrated on Gamma Knife® Icon™. *J Radiosurg SBRT*. 2019;6(2):121-129. http://www.ncbi.nlm.nih.gov/pubmed/31641548
12. Hubley E, Mooney KE, Schelin M, Shi W, Yu Y, Liu H. Geometric and dosimetric effects of image co-registration workflows for Gamma Knife frameless radiosurgery. *J Radiosurg SBRT*. 2020;7(1):47-55. http://www.ncbi.nlm.nih.gov/pubmed/32802578
13. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. *IEEE Trans Med Imaging*. 1997;16(2):187-198. https://doi.org/10.1109/42.563664
14. Studholme C, Hill DLG, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit*. 1999;32(1):71-86. https://doi.org/10.1016/S0031-3203(98)00091-0
15. Wang Z, Bovik AC. Mean squared error: love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Process Mag*. 2009;26(1):98-117. https://doi.org/10.1109/MSP.2008.930649
16. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13(4):600-612. https://doi.org/10.1109/TIP.2003.819861
17. Premium System Phantom. qmri. Accessed February 8, 2021. https://qmri.com/qmri-solutions/t1-t2-pd-imaging-phantom/
18. Magphan System Phantom S162. The Phantom Laboratory. Accessed February 8, 2021. https://www.phantomlab.com/magphan-s162
19. System Phantom. NIST/ISMRM. Accessed February 8, 2021. https://www.nist.gov/programs-projects/quantitative-mri
20. FUNSTAR. Gold Standard Phantoms. Accessed February 8, 2021. https://www.goldstandardphantoms.com/funstar
21. SPECTRE. Gold Standard Phantoms. Accessed February 8, 2021. https://www.goldstandardphantoms.com/spectre
22. Diffusion Phantom. qmri. Accessed February 8, 2021. https://qmri.com/qmri-solutions/adc-dwi-imaging-phantom/
23. Diffusion Phantom. NIST/RSNA/NCI. Accessed February 8, 2021. https://www.nist.gov/programs-projects/quantitative-mri
24. SRT (Steriotactic Radiosurgery) LiquiPhil End-To-End Phantoms. The Phantom Laboratory. Accessed February 8, 2021. https://www.phantomlab.com/srt-phantoms

25. Phantom Patient for Stereotactic End-To-End Verification. CIRS. Accessed February 8, 2021. https://www.cirsinc.com/products/radiation-therapy/phantom-patient-for-stereotactic-end-to-end-verification/
26. MR Distortion & Image Fusion Head Phantom. CIRS. Accessed February 8, 2021. https://www.cirsinc.com/products/radiation-therapy/mr-distortion-image-fusion-head-phantom/
27. Adult Head (Standard). True Phantom Solutions. Accessed February 8, 2021. https://truephantom.com/product/adult-head-standard/
28. Adult Head (Static). True Phantom Solutions. Accessed February 8, 2021. https://truephantom.com/product/static-adult-head-for-rd/
29. Head Phantom. PureImaging Phantoms. Accessed February 8, 2021. https://www.pureimagingphantoms.com/product/head-phantom/
30. Fischer H, Ladebeck R. *Echo-planar imaging image artifacts.* Echo-Planar Imaging. Springer; 1998:179-200. https://doi.org/10.1007/978-3-642-80443-4_6
31. Jungmann PM, Agten CA, Pfirrmann CW, Sutter R. Advances in MRI around metal. *J Magn Reson Imaging*. 2017;46(4):972-991. https://doi.org/10.1002/jmri.25708
32. Greitz D, Wirestam R, Franck A, Nordell B, Thomsen C, Ståhlberg F. Pulsatile brain movement and associated hydrodynamics studied by magnetic resonance phase imaging. *Neuroradiology*. 1992;34(5):370-380. https://doi.org/10.1007/BF00596493
33. Gomes JGR, Gorgulho AA, de Oliveira López A, et al. The role of diffusion tensor imaging tractography for Gamma Knife thalamotomy planning. *J Neurosurg*. 2016;125(Supplement_1):129-138. https://doi.org/10.3171/2016.7.GKS161553
34. Mehrabian H, Myrehaug S, Soliman H, Sahgal A, Stanisz GJ. Quantitative magnetization transfer in monitoring glioblastoma (GBM) response to therapy. *Sci Rep*. 2018;8(1):2475. https://doi.org/10.1038/s41598-018-20624-6
35. Mehrabian H, Myrehaug S, Soliman H, Sahgal A, Stanisz GJ. Evaluation of glioblastoma response to therapy with chemical exchange saturation transfer. *Int J Radiat Oncol*. 2018;101(3):713-723. https://doi.org/10.1016/j.ijrobp.2018.03.057
36. Chan AY, Tran DKT, Gill AS, Hsu FPK, Vadera S. Stereotactic robot-assisted MRI-guided laser thermal ablation of radiation necrosis in the posterior cranial fossa: technical note. *Neurosurg Focus*. 2016;41(4):E5. https://doi.org/10.3171/2016.4.FOCUS1622
37. Shen G, Xu L, Xu M, Geng M, Tan Y, Li F. 1H-MR spectroscopy guided Gamma Knife radiosurgery for treatment of glioma. *Turk Neurosurg*. 2012;22(6):690-694. https://doi.org/10.5137/1019-5149.JTN.5676-12.1

How to cite this article: Oglesby RT, Lam WW, Ruschin M, et al. Skull phantom-based methodology to validate MRI co-registration accuracy for Gamma Knife radiosurgery. *Med Phys*. 2022;49:7071–7084. https://doi.org/10.1002/mp.15851