2019

Effects of reference image selection on the alignment of free-breathing lung cancer patients during setup imaging: average intensity projection versus mid-ventilation

Samantha Conrad

Effects of reference image selection on the alignment of free-breathing lung cancer patients during setup imaging: average intensity projection versus mid-ventilation

A thesis submitted in partial fulfillment of the requirements for the Degree of Master of Science in Medical Physics at Virginia Commonwealth University

By

Samantha Gail Conrad

Bachelor of Science in Physics, University of Kansas, May 2017

Advisor: Dr. Laura Padilla

Assistant Professor, VCU Department of Radiation Oncology, Division Medical Physics

Virginia Commonwealth University

Richmond, Virginia

May 2019

ACKNOWLEDGEMENTS

I would like to thank and acknowledge my committee members, Dr. Elizabeth Weiss, Dr. Frank Corwin, and Dr. Mihaela Rosu, and especially my advisor Dr. Laura Padilla. I would also like to thank those who volunteered their time for registrations. Finally, I would like to thank my peers and my partner, Zack Hames, for their continued support.

Contents

List of Tables

Appendix Tables

List of Figures

Main Text

Appendix Figures

Introduction

Lung Cancer, 4DCT, and Free-Breathing Conditions

The American Cancer Society estimates that 1 in 56 people will develop lung or bronchial cancer in their lifetime [1]. Furthermore, lung cancer is the leading cause of cancer death [1]. These two factors make treating lung cancer a relatively common and important practice.

The general workflow for radiotherapy treatments includes 3 steps: simulation, treatment planning, and treatment delivery. During simulation, the patient's treatment position and immobilization devices are selected, and a planning CT is acquired. This CT is used to design the treatment for the patient to the fulfill the physician's requested prescription. The treatment plan is done based on the normal organs and target structures contoured on the images. Finally, treatment delivery is performed based on the approved treatment plan. For more complex and precise treatments, image guided treatment delivery is performed through the use of cone-beam computed tomography (CBCT) [15][16][17]. When treating lung cancer patients, one of the many obstacles faced by clinicians is breathing motion [2]. Radiation oncology clinics are able to account for this phenomenon in several ways [3]. Many lung cancer patients undergo a 4D Computed Tomography Simulation (4DCT Sim) under free-breathing (FB) conditions [12][13][14]. The 4DCT sorting is phase-based which means that the software divides the breathing cycle in equal time intervals. A respiratory motion tracker provides the software with the patient's breathing signal as the scan is being acquired [13][27]. Through this information, the software categorizes and assigns the projections acquired during those times into their respective bins [13]. The resulting information obtained from the 4DCT consists of 10 breathing cycle "phases" which span from full inhale (0%) through full exhale (50%) to near full inhale (90%). From these phases, several "reference images" can be obtained [11]. Reference images are used for contouring the internal target volume (ITV) and organs at risk (OARs) [9], treatment planning/dose calculation, and for patient alignment prior to treatment delivery.

A reference image that has a high fidelity to its intended purpose and expected qualities is, of course, ideal. Simply put, proper treatment requires the reference image to reflect the truth. If for any reason the reference image does not accurately represent the patient's internal anatomy and extent of motion, inappropriate treatment planning and patient alignment during treatment may occur.

Typically, the ITV is contoured using all 10 phases obtained from the 4DCT or it is obtained by the Maximum Intensity Projection (MIP) of that 4DCT [6][9]. In general, the same reference image used for contouring OARs and treatment planning is used for patient alignment at the machine. Therefore, this reference image is also required to accurately reflect the tumor throughout the breathing cycle which can be easily compared to the CBCT taken for patient alignment.

The goal of this study is to help determine which reference image, MidV or AIP, is best for patient alignment during free breathing treatments. Ultimately, the reference image will need to be registered with the CBCT obtained before treatment for patient positioning. A proper comparison of how the selected reference images affect registrations when registered to CBCTs is necessary. Several of these comparisons have been performed in recent years [18][19][20][21][22][23][24][25]; however, they have either compared different reference images than addressed in this paper or have not examined the effects of various common issues faced with lung cancer patient treatment (i.e. high tumor motion, breathing cycle variability, etc.) on image registration for patient alignment that this paper attempts to address. A summary of papers that compared different reference images for patient alignment purposes can be found in the Previous Works section of this thesis.

30% Phase Mid-Ventilation Surrogate

The reference image of choice at VCU is the 30% phase (MidV) image obtained from the 4DCT. This phase is chosen as the mid-ventilation surrogate because more time is spent near exhalation rather than inhalation during normal respiration, hence the 30% phase is thought to better represent the mid-ventilation point. The result of using this reference image is a visually sharp “snapshot” of the tumor and surrounding organs at risk during its motion cycle [29].

This reference image can sometimes have high motion artifacts associated with it due to the high motion gradient of the breathing cycle at the 30% phase [28][31]. These artifacts can change the visual shape and size of the tumor [31]. Furthermore, assuming that the 30% phase is representative of the mid-ventilation point or time-averaged mean tumor position is a broad assumption that may or may not apply to every patient during every simulation and/or treatment. It is well documented that breathing cycles vary between individuals and between treatments [2][30]. If the MidV reference image is not accurate to the mean tumor location during breathing

or the breathing during simulation and treatment are different, this could result in a systematic offset in alignment when using this reference image. The offset could potentially lead to under-treating the tumor as well as over-dosing the surrounding OARs.

Average Intensity Projection

Another reference image option under consideration by VCU and in current practice at many clinics is the use of the Average Intensity Projection (AIP). This image is created by averaging the intensity of the images over the10 phases of the breathing cycle obtained by the 4DCT; this average represents the tumor's location throughout the respiratory cycle. The result of using this image is having an image with a blurred area around the tumor's path with a brighter inside where the tumor spends more time over all the breathing cycles captured by the scan.

This reference image is not as susceptible to artifacts and its selection does not make any specific assumptions about the patient's breathing cycle; it depicts the full extent of motion of the tumor throughout respiration during simulation. Theoretically, it should be much closer in nature to the CBCT obtained before treatment since both images represent the tumor position over the whole breathing cycle. This means that it could be a better choice for patient alignment. However, any reference image obtained under FB conditions assumes that the breathing cycle is consistent between simulation and treatment, which may not be true [8]. Furthermore, it can be difficult to align a motion-blurred AIP image with a lesser quality CBCT image with confidence at the treatment machine. For this reason, the AIP might not be the most clinically effective reference image.

Brief Comparison

Each reference image has their own set of pros and cons. The MidV method makes the broadest assumptions (30% phase is representative of the mid-ventilation point); they are assumptions that the use of the AIP effectively eliminates. Although most patients may spend more of their breathing cycles in the exhale position, every patient has the full capability to follow a different pattern, especially under FB conditions. This can affect the phase that most accurately represents the MidV position. The AIP makes no assumptions on the mean tumor position and reports the full range of motion of the tumor. The CBCT obtained at the machine

similarly shows a blurred track of where the tumor is during acquisition. Again, this means that, in theory, the AIP and the CBCT have more in common visually and both can represent the range of locations of the tumor during their respective acquisitions. However, the difficulty of registering two blurry images may at times prove difficult and may be too subjective in certain cases. Moreover, even if both the MidV and AIP reference images are accurate to their definitions, it cannot be assumed that the breathing cycle at simulation will match the breathing cycle during treatment.

The most ideal reference image must have faithfulness to its intended definition, consistency amongst various observers, and consistency from the same observer even under variable breathing cycles. It is not likely that one method will have complete immunity to the effects of common issues with lung cancer patients, but the question whether one is more reliable remains. This paper's objective is to compare the inter-observer and intra-observer reproducibility of alignments and investigate how physical characteristics (rotation, tumor size, etc.) or breathing cycle changes can affect these two alignment methods.

Indicators of Quality for Reference Images

In order to establish a quality method of alignment, a few important indicators would be desired. The first one, and the most difficult to assess, would be accuracy. In short, the clinic needs to choose a reference image that can best help in aligning the patient so that the treatment delivered most closely matches what was planned. Both the MidV and AIP reference image methods are based off assumptions. The 30% phase MidV reference image assumes 30% phase is a good indicator of the mean position of the tumor during the respiratory cycle. When treating under free-breathing conditions, as we do at VCU, both the MidV and AIP reference image methods assume that the breathing cycle is consistent between the acquisition of the 4DCT scan and all treatment fractions. It is regularly observed in the clinic and has been shown in research not to be the case [2]. Furthermore, these breathing cycle variations not only can impact patient alignment but can ultimately affect the fidelity of the ITV contour (subsequently the PTV) used for treatment planning.

Another indicator of a quality alignment method would be the reproducibility and consistency for an individual's alignments. This is generally important because we want the

same observer (i.e. the therapist, or radiation oncologist) to have a high level of confidence in being able to locate and align the tumor.

A third indicator of quality is the reproducibility and consistency of alignments amongst different individuals. This is important because the patient alignment for treatment should be reliable and consistent regardless of who is performing the image registration. Intuitively, higher variability between observers is expected due to differences in individual perspectives and interpretation of visual information. Ideally, the reference image would have visual characteristics that are unambiguous so that any given individual can reach an alignment that is within the bounds of the PTV margin [4].

In this analysis, a method fails for a patient if any fraction shows variability amongst an individual or in a group that exceeds the clinical PTV margin [27] At VCU, margins typically are 5 mm or higher. If an alignment method does not have these three indicators of quality, then it its clinical use could compromise the quality of treatment. If the two methods (MidV or AIP) meet these requirements, the clinic can choose either the MidV or AIP reference image for their practice.

In addition to the three above indicators, a more favorable method would include inherent immunity to common issues seen in lung cancer patients. These issues can include patient rotation, tumor size, breathing cycle variability (when treating under FB conditions), tumor motion, regression of tumor, or artifacts. The small sample size of this paper's population does not allow for firm statistical evaluation of all issues correlating to alignment methods; however, the next subsection describes the theoretical impact these physical characteristics can have.

Physical Characteristics

Since patient setup rotation is only manually estimated and corrected inside the treatment room, this can have a large effect on how different individuals align the setup and reference images. When two objects must be registered together but one is rotated, it is subjective where the best alignment is based solely on translations. These variations can be amplified when observers align focusing on different portions of the tumor. One observer may align about their perceived axis of rotation, one observer may pick the edge closest to the chest wall, or one observer may pick the largest tumor feature to align to. Even if philosophies are the same for

observers and shifts are performed for the same slices of the tumor, it is still a matter of perspective where the “correct” shift lies.

Tumor motion is an inevitable barrier when treating under free-breathing conditions. From the ITV to treatment alignment, it is an ever-present physical characteristic shared by all patients in various amplitudes. Tumor motion is responsible for artifacts and location uncertainties at all times. These issues can theoretically play a major role in the failure of quality registrations. The presence of artifacts in either the simulation data or the CBCT’s acquired at treatment stand as visual obstacles that can disrupt a proper alignment. When present, special care must be taken during alignment. Otherwise, the registrations may result in any of the three quality failures discussed.

The presence of artifacts in either the simulation data or the CBCTs acquired at treatment stand as visual obstacles that can disrupt a proper alignment. When present, registrations may result in any of the three quality failures discussed.

Tumor growth or regression, and other anatomical changes, such as pleural effusion and atelectasis, were not quantified in this paper but can theoretically impact a group’s ability to align within the proper agreement. In future studies, this may be investigated.

Another physical quality that can lead to registration failure is breathing cycle changes between simulation and subsequent treatments. These types of changes can affect the appearance of the tumor in the CBCT relative to the reference images, and therefore, this can introduce difficulty in registering two images that are not visually similar.

Previous Works

Below is a review and summary of research papers that investigated and compared various reference images.

Jiang et al. [18] compared setup errors (translation and rotational) obtained with CBCT with free-breathing CT images and AIPs. They explored the correlation between the difference of translational setup errors and the gross tumor volume (GTV) motion. The setup errors were determined based on the image registration between the CBCT and the 2 different reference images, respectively. Jiang et al. found that the translational setup errors based on FB-CT were significantly larger than those from AIP in the left/right, superior-inferior, and anterior-posterior directions. They concluded that the AIP should be the preferred choice of reference images when compared to a FB CT.

Shirai et al. [19] determined whether MIP or AIP is preferred for alignment to CBCT images. Using a phantom with no respiratory motion at first, stationary CBCT was aligned to stationary CT, and the couch position was used as the baseline. A CBCT image of a phantom mimicking breathing motion was aligned to the MIP and AIP of corresponding amplitudes. Registration error was defined as the SI deviation of the couch position from the baseline. In addition, using patient data the free-breathing CBCT was registered to AIP and MIP and the difference in couch shifts was calculated. Shirai et al. concluded that the AIP is recommended as the reference image for registration to FB-CBCT; MIP causes systematic target positioning error.

Komomae et al. [20] assessed the accuracy of target position localization using free-breathing CBCT during stereotactic lung radiotherapy while using the AIP for registration. This group used a known respiratory motion phantom. They demonstrated a sufficient level of geometric accuracy (1.07 ± 1.23 mm) using the free-breathing CBCT and the image-guidance system mounted on the Vero4DRT. However, the inter-observer variation and systematic localization error of image matching substantially affected the overall geometric accuracy.

Oechsner et al. [21] analyzed setup errors resulting from image registration of slow planning CT (PCT), AIP, MIP, and MidV with free breathing cone beam CTs (FB-CBCT). The results were correlated to tumor characteristics. Oechsner concluded that the AIP CTs yielded the smallest shift differences and might be the most appropriate CT dataset for registration with 3D FB-CBCTs.

Oechsner et al. [22] assessed the impact of slow planning CT (PCT), AIP, MIP, and MidV manual image registration with FB-CBCT for patient positioning by several observers. This group found that the mean 3D shift difference between different reference CT datasets was the smallest for AIP vs MIP and the largest for MidV vs PCT (differences >10 mm). The smallest differences were found for AIP, which might be the most appropriate CT dataset for image registration with FB-CBCT.

Hugo et al. [23] compared two 4D methods (AIP/CBCT and RCCT/RC-CBCT) for image-guided target localization in the presence of respiration. A respiration correlated CT (RCCT) was acquired on a CT Simulator, and an AIP image was generated from the RCCT. A respiration-correlated cone-beam CT (RC-CBCT) and a free-breathing cone beam CT (FB-CBCT) were acquired. The ‘‘RCCT method’’ consisted of calculating the mean target position on both the RCCT and RC-CBCT, registering the RCCT to the RC-CBCT, and determining the shift in the mean target position from the planned mean position. The ‘‘AVG-CT method’’ consisted of registering the AIP to the FB-CBCT. The ability of each to measure the shift in the mean target position was compared, both in a respiratory phantom and in 8 patients. Hugo et al. concluded that the RCCT method enables localization of the mean tumor position and measurement of changes in the motion pattern, whereas the AIP method is simple, fast, and easily implemented.

Shirai et al. [24] compared registration variations between MIP-CBCT and AIP-CBCT. This group concluded that in the phantom study, the couch position at AIP registration was close to the baseline. In both phantom and patient study, the couch position using the MIP registration was worse than the AIP registration. Therefore, AIP registration is more appropriate.

Oechsnet et al. [25] evaluated the differences of using slow planning CT (PCT), AIP, MIP, and MidV CT datasets for image registration with free breathing cone beam CTs (CBCT). This group concluded that using different CT datasets for image registration with free breathing CBCTs can result in different couch shifts. Automatic AIP-CBCT and MIP-CBCT had the best agreement.

This thesis differs from previous works in that it combines both manual intra-observer registration variability and manual inter-observer registration variability when using either the AIP or 30% phase reference images from patients, retrospectively. Furthermore, this thesis

quantifies several physical characteristics and analyzes their effects on registration variability. Some physical characteristics have been analyzed and compared previously (tumor motion); this thesis is the first to attempt to correlate rotation, tumor size, and breathing motion changes with registration variability.

Statement of Problem

The purpose of this paper is to quantify if using an average intensity projection (AIP) scan or a 30% phase (mid-ventilation, MidV) scan as the reference image affects reproducibility of lung cancer patient alignment under free-breathing cone beam computed tomography (CBCT) image guidance and analyze the effects of common clinical issues on these results.

Methods

This study was reviewed and deemed IRB exempt by the review board at Virginia Commonwealth University (study number: HM20014421). A total of 16 lung cancer patients (14 SBRT, 2 conventional) with all 10 breathing phases from their 4DCT were retrospectively investigated. All 16 patients were originally planned and treated using the 30% phase reference image. A radiation oncologist physician contoured the OARs and the ITV. ITV contours were contoured using the MIP image from each patient's 4DCT. AIPs were retrospectively created from the 4DCT data. Each of the patients had between 3-5 fractions correlating to 3-5 CBCTs for alignment.

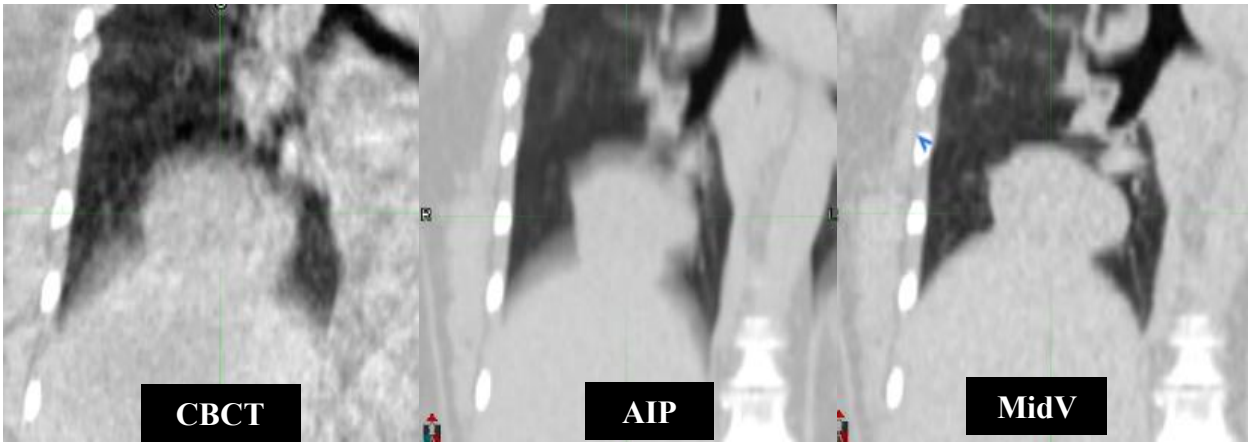

Figure 1. Shows the cone-beam computer tomography (CBCT), average intensity projection (AIP) reference image, and 30% phase mid-ventilation (MidV) reference image of the same tumor. AIP and MidV images are from the same slice. Magnification of all three images is consistent. Visually demonstrates blurred nature of CBCT images, blurred nature of AIP images, and sharp nature of MidV images. The AIP was registered with the CBCT, and the MidV was registered with the CBCT.

Using MIM 6.2 (Beachwood, OH) software, manual, rigid registrations were performed between the AIP and each CBCT (AIP/CBCT) and between the MidV and each CBCT (MidV/CBCT). The registrations were performed by 5 individuals (medical physics student, medical physics resident, radiation oncology resident, medical physicist, and radiation oncology attending physician). The "Lung" Window/Level setting for all images was used by all observers. The images were rigidly registered, ITV contours were turned on, and no rotations were allowed in order to reflect real treatment conditions at VCU. All observers were asked to register the images to align the tumor as best as possible overall based on its shape, salient features when applicable, and extent of motion. The alignment ultimately fell to the observer's discretions, which is an important part of this study.

Once the observer obtained an acceptable registration, the corresponding registration shifts were recorded for the lateral, anterior-posterior, and superior-inferior directions.

To avoid biasing, observers were asked to not compare between their MidV/CBCT and AIP/CBCT shifts.

Additionally, the AIP/CBCT registrations, the MidV/CBCT registrations, and the MIP/PMIP were repeated 3 times by one individual for intra-observer variability assessment. To avoid biasing, each fraction was completed once using the same reference image until all fractions were completed. After all registrations had been done once with one reference image, the process was repeated using another reference image. Several weeks were spent between the three trials to further avoid any biasing.

Rotations were estimated using an automatic rigid box-alignment registration (between MidV/CBCT and between AIP/CBCT) in MIM. Due to time constraints, only the two fractions with the highest average standard deviation (1 MidV and 1 AIP) per patient were selected. For some patients, this was the same fraction. The rigid box-alignment region of interest was placed to encompass the tumor and ITV contour. MIM software automatically registered the two images with translational shifts (LAT, AP, and SI) and rotations (yaw, pitch, roll). If both registrations were from the same fraction, their individual angles were still investigated and recorded, and their results were checked to verify the registrations resulted in similar angles of rotation for both reference images. All box-alignments were checked to confirm a quality registration.

Tumor volume was estimated by using the ITV contour volume, so it represents both tumor volume and extent of motion. Tumor motion was estimated by placing a virtual grid over each patients 4DCT and estimating the distance the tumor traveled in all three directions throughout the full breathing cycle.

In an effort to estimate the effects of breathing cycle variations between simulation and each treatment, another registration was performed as a baseline. A pseudo-MIP (PMIP) was created for each treatment fraction by adjusting the Window/Level of the CBCT to brightly show the tumor's location throughout the acquisition. Since exhale has been shown to be the most consistent position of the respiratory cycle [7], the most superior edge of the tumor (exhale position) of the PMIP was registered to the most superior edge of the simulation MIP. If the amplitude of the breathing cycle between that treatment fraction and the simulation were similar, the registration of the CBCT to the AIP and this baseline registration should match. Hence,

taking the difference between PMIP/MIP registration and the ones with the standard image sets should, in theory, be a reasonable surrogate to estimate breathing cycle variations.

Results

Averaged Population Results

All registrations resulted in a recorded shift (units: mm) for the lateral (LAT_{MidV} and LAT_{AIP}), anterior-posterior (AP_{MidV} and AP_{AIP}), and superior-inferior (SIM_{idV} and SI_{AIP}) directions for each fraction.

The average standard deviations in the lateral (LAT), anterior-posterior (AP), and superior-inferior (SI) directions for the intra-observer registrations of the MIP/PMIP set, AIP/CBCT set, and MidV/CBCT set are calculated and reported in Table 1. The magnitude of the average intra-observer registrations is also reported (Eq. 1). QS StDev is the summation in quadrature of all three directions' (LAT, AP, and SI) standard deviations. The average standard deviations in the LAT, AP, and SI directions for the inter-observer registrations (containing the average of the three sets from the intra-observer registrations) of the AIP/CBCT sets and MidV/CBCT sets are also calculated.

$$QS\ StDev = \sqrt{(StDev_{LAT})^2 + (StDev_{AP})^2 + (StDev_{SI})^2} \quad (1)$$

Table 1. Table showing all averaged results for the 16-patient population. Includes the average standard deviations of AIP/CBCT and MidV/CBCT intra-observer registrations and their respective summations in quadrature (Eq. 1). Also includes the average standard deviations of the AIP/CBCT and MidV/CBCT inter-observer registrations and their respective magnitudes (Eq. 1). All average standard deviations are reported in the lateral (LAT), anterior-posterior (AP), and superior-inferior (SI) directions. Units are in mm.

	LAT (mm)	AP (mm)	SI (mm)	QS StDev (mm)
AIP/CBCT Intra	**0.4**	**0.4**	**0.7**	**0.9**
MidV/CBCT Intra	**0.4**	**0.7**	**0.9**	**1.2**
AIP/CBCT Inter	**0.7**	**0.8**	**1.3**	**1.7**
MidV/CBCT Inter	**0.7**	**1.0**	**1.3**	**1.8**

Per-Patient and Per-Fraction Registration Results

Comparing Individuals' MidV/CBCT and AIP/CBCT Registrations

The differences between each observer's MidV/CBCT registration and AIP/CBCT registration (LAT, AP, and SI directions) was calculated for each fraction. The magnitude (ΔMidV&AIP) of this difference was calculated per fraction (Eq. 2). The average intra-observer registration and the other four observer's registrations were averaged together for both reference images. Figures 2 and 3 show the highest value found by Eq. 2 (black lines), the difference between the averaged MidV and average AIP registrations (blues line), and the average value found by Eq. 2 (green lines). No results were found to exceed a 10 mm difference between the MidV and AIP registrations.

$$\Delta MidV\&AIP = \sqrt{(LAT_{MidV} - LAT_{AIP})^2 + (AP_{MidV} - AP_{AIP})^2 + (SI_{MidV} - SI_{AIP})^2} \quad (2)$$

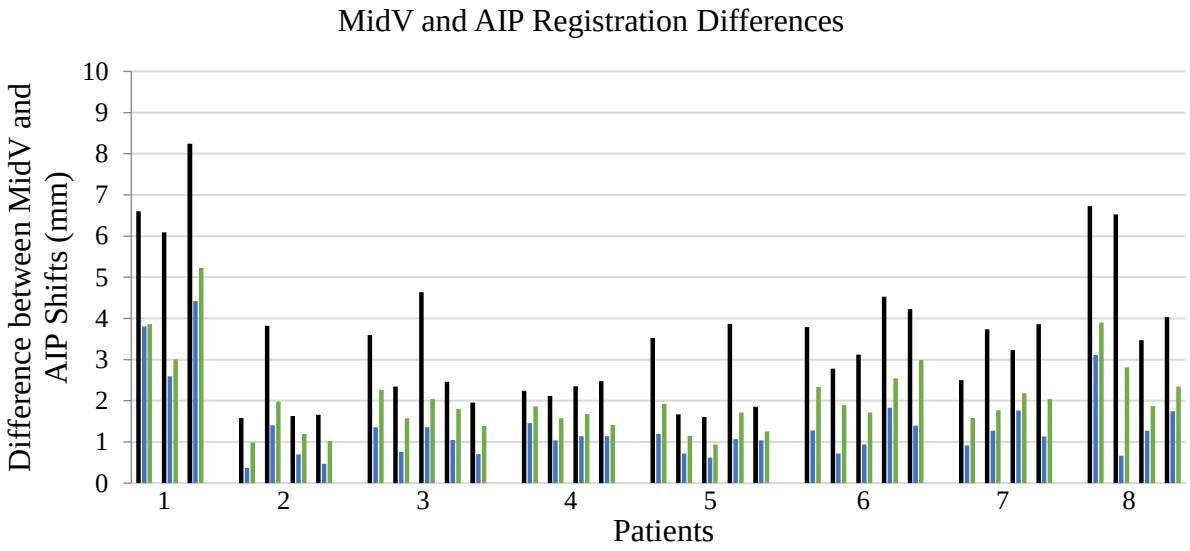

Figure 2. Graph of the maximum difference between MidV/CBCT and AIP/CBCT registrations done by the same observer (Eq. 2, black lines), the difference between the average MidV registration and average AIP registration (blue lines), and the average difference between observer's MidV and AIP registration (green line) per fraction. Patients 1-8 (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background. Patients 1 and 8 have at least a single fraction with a difference between the MidV/CBCT and AIP/CBCT registrations done by the same observer above 6 mm.

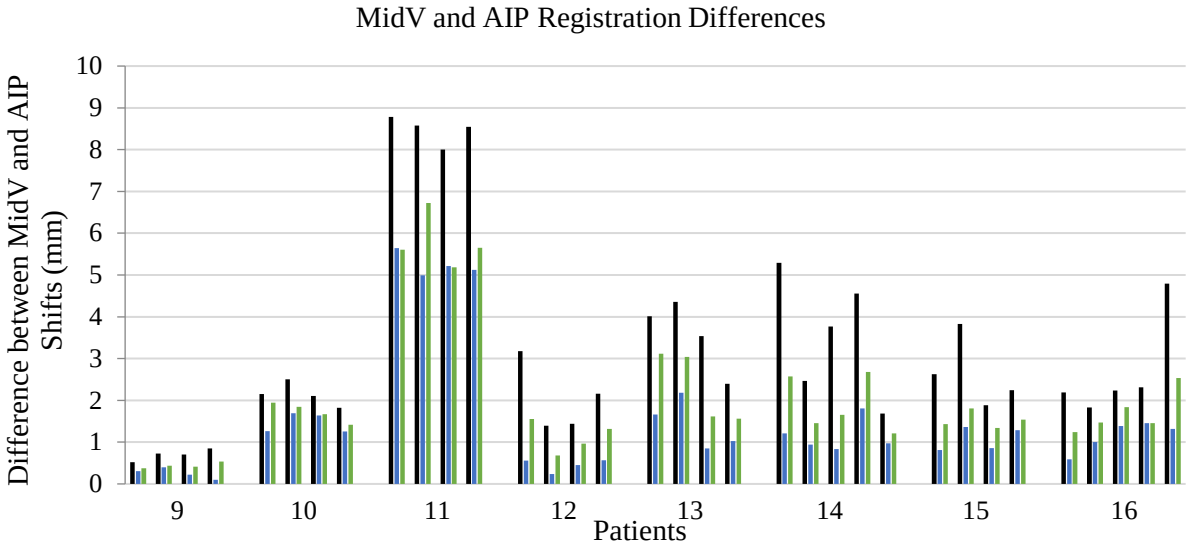

Figure 3. Graph of the maximum difference between MidV/CBCT and AIP/CBCT registrations done by the same observer per fraction (Eq. 2, black lines), the difference between the average MidV registration and average AIP registration (blue lines), and the average difference between observer's MidV and AIP registration (green line). Patients 9-16 (fractions ranging from 4-5) are shown; patient 9 is the left most set of fractions (white background, four fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background. Patient 11 had several fractions with a difference between the MidV/CBCT and AIP/CBCT registrations done by the same observer above 6 mm.

Inter-Observer Results

From this point on, the MidV/CBCT registrations will be denoted by "MidV registrations"; the AIP/CBCT registrations will be denoted by "AIP registrations."

All combinations of the difference between inter-observer registrations (LAT, AP, and SI directions) for each reference image set respectively, per fraction, were calculated. The magnitude of those differences was taken to determine the magnitude of inter-observer variation (Eq. 3). ΔIntObs. is the magnitude of the difference between the registrations of two observers. $\Delta LAT_{n,m}$ is the difference between the lateral registration shifts of observer n, and observer m, for the same fraction. The same logic follows for the anterior-posterior (AP) and superior-inferior (SI) directions.

$$\Delta Inter\ Obs.MidV\ or\ AIP = \sqrt{(\Delta LAT_{n,m})^2 + (\Delta AP_{n,m})^2 + (\Delta SI_{n,m})^2} \quad (3)$$

The average difference between inter-observer registrations using either reference image was also calculated for comparison.

Figure 4 shows the difference in inter-observer variabilities between the MidV data set and AIP data set. Black lines represent the difference in maximum inter-observer registration variability. Blue lines represent the difference in the average registration difference within the inter-observer data set. A positive value indicates that the MidV variability was higher than the AIP; a negative value indicates the AIP variability was higher than the MidV.

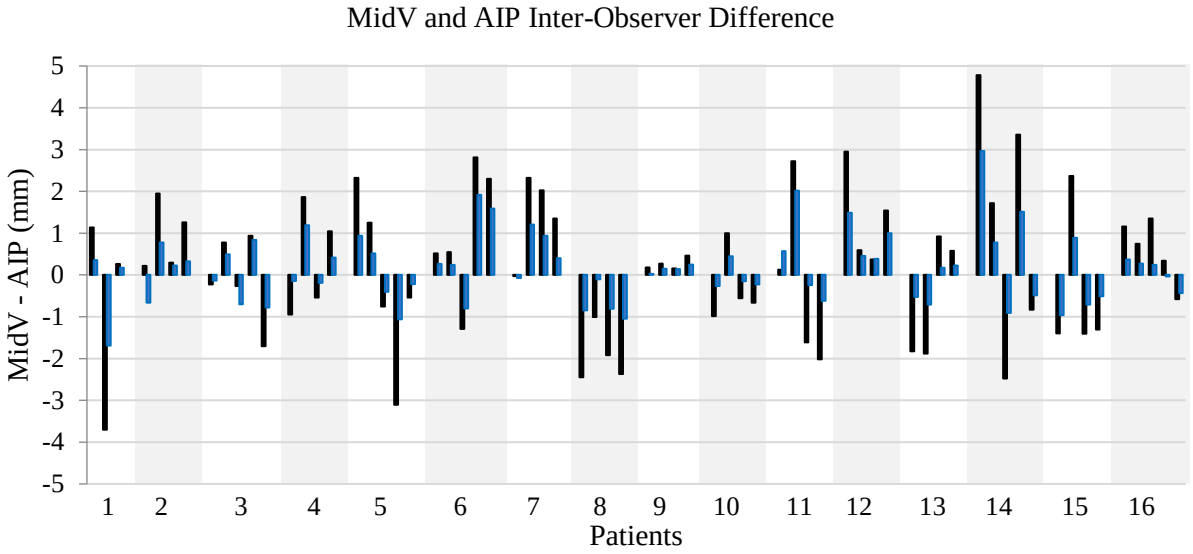

Figure 4. Shows difference between the highest inter-observer variability of the MidV and AIP data sets and the difference between the average MidV difference and AIP difference. A positive number indicated the inter-observer variability for the MidV data set was higher than that of the AIP. A number close to zero shows that the variability between the two reference images was similar. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background.

Intra-Observer Results

Intra-observer variations were analyzed following the same procedure for inter-observer registrations; results showed that these values were on average 1 mm smaller than inter-observer variability and hence will not drive the selection of reference image. Average standard deviations of the intra-observer variability were always less than that of inter-observer variability. The purpose of this paper is to determine which reference image, if either, is most suitable for the clinic. Since results were less than or equal to inter-observer results, they have not been reported in the main text but are reported in the Appendix.

Physical Characteristic and Common Issues

Table 2 contains compiled physical characteristics that were more closely examined to see if their presence gave a propensity for registration uncertainties. These physical characteristics are rotation in the CBCT, tumor volume, tumor motion, and tumor location. Breathing cycle variability is reported later.

The highest rotation of the two registrations is reported in Table 2. The volume of the ITV contour for each patient is also reported. The apparent tumor motion in the LAT, AP, and SI directions based on the 4DCT data is reported alongside the magnitude of all three directions and the lung region.

Table 2. Table showing the complied physical characteristics for all 16 patients relating to some common issues had when treating lung cancer. The two registrations (1 MidV and 1 AIP) with the highest inter-observer standard deviation were selected for box-alignments to confirm whether rotation was present. The highest angle rotation is recorded. Both box-alignments were ensured to have similar rotations and be quality alignments. Table also shows the volume of the ITV contour. The averaged tumor motion in the lateral (LAT), anterior-posterior (AP), and superior-inferior (SI) directions are recorded as well as the magnitude (QS) of the three directions and the tumor's lung region.

	Largest MidV | ∠ |	Largest AIP | ∠ |	ITV Vol (cc)	Tumor Motion (cm)				Region
				LAT	AP	SI	QS	
1	2°	2°	90.6	0	0.5	1.3	1.4	RLL
2	3°	2°	5.3	0	0.5	0.75	0.9	LUL
3	5°	5°	45.4	0.25	0.75	1.25	1.5	LLL
4	1°	1°	9.5	0.75	0.25	0.75	1.1	LUL
5	1°	2°	5.59	0.5	0	0.5	0.7	RUL
6	8°	8°	15.5	0.25	0.5	0.75	0.9	RLL
7	4°	4°	22.1	0	0	1	1.0	RLL
8	1°	1°	2.1	0.25	1.25	1.3	1.8	RUL
9	1°	1°	1.0	0	0	0.25	0.3	RUL
10	1°	1°	2.3	0	0.5	0.5	0.7	RUL
11	10°	10°	26.1	0.5	0.5	1.25	1.4	LLL
12	2°	1°	7.0	0	0.5	0.25	0.6	LUL
13	4°	4°	166.5	0	0.5	1	1.1	RLL
14	2°	2°	111.7	0	0	0.75	0.8	RLL
15	1°	1°	33.6	0	0.25	1	1.0	LLL
16	3°	3°	50.1	0	0	0.25	0.3	Carina

Figure 5 shows the correlation of rotation and differences between inter-observer registrations for the MidV and AIP data sets, respectively. In other words, it shows the effect of rotation on the inter-observer reproducibility. The maximum inter-observer variability used was only from the same fraction which the rotation angles were found.

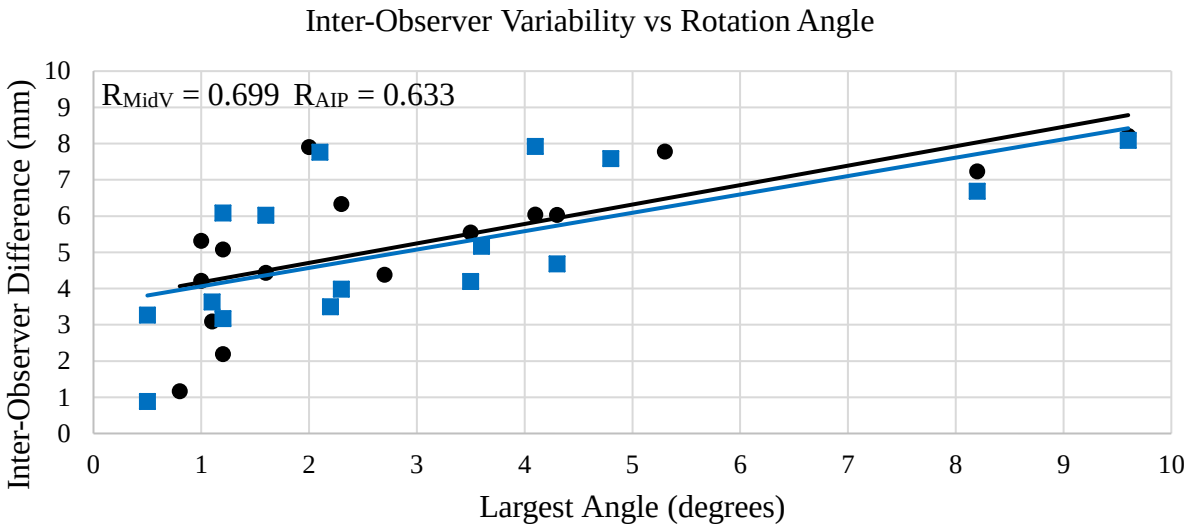

Figure 5. Scatter plot of largest rotation angle (Table 2) vs inter-observer variability (Eq. 3) for the MidV fraction with the highest int. obs. standard deviation (black circle) and the AIP fraction with the highest inter-observer variability (blue squares). Can show if inter-obs. variability is increased when the patient is rotated. Shows positive trend for both reference images. No points with a rotation angle > 4° had a variability below 6 mm using the MidV, one point had a variability less than 6 mm with a rotation above 4° using the AIP. R values = 0.699 and 0.633 for MidV and AIP, respectively.

Figure 6 shows the correlation between the tumor motion and the differences between the maximum inter-observer registrations for every fraction in the MidV and AIP data sets, respectively. These plots show if a tumor with higher motion results in wider inter-observer variability.

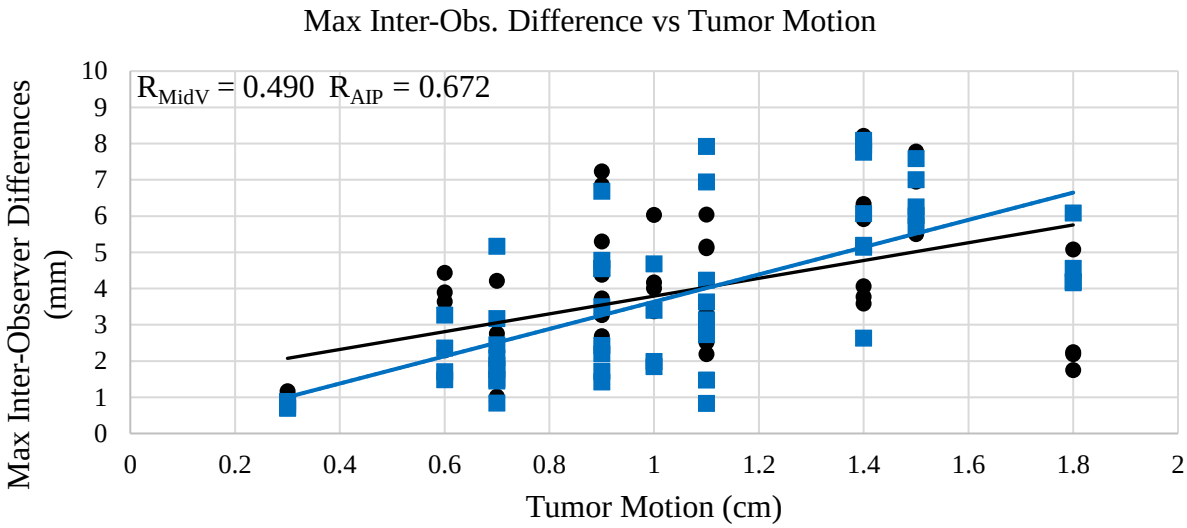

Figure 6. Scatter plot of magnitude of the estimated tumor motion (Table 2) vs the highest MidV (black circles) or AIP (blue squares) inter-observer variability (Eq. 3) for the all fractions. Can indicate if/how inter-obs. variability can be affected by tumor motion. Shows positive trend for both reference images and comparable correlations. All points with a variability > 6 mm have tumor motion > 0.8 mm. R values = 0.490 and 0.672 for MidV and AIP, respectively.

Correlation between the uncertainty in the registrations and breathing cycle variations (as estimated with the PMIP/MIP methodology) did not show conclusive results. Detailed results of this study can be found in the Appendix.

Discussion

Causes for Registration Variability

The determination of alignment accuracy (the selected patient alignment leads to a treatment delivery that matches what was planned) is outside of the scope of the methodology presented by this paper. In order to determine if an alignment is accurate, a reliable, objective, "ground truth" alignment is necessary for comparison. No such information is available at this time for our patient cohort. Accuracy may be compared in future studies with the use of a phantom; for this retrospective patient study, this is not possible. For this reason, registration variability is used to compare the quality of the two available reference images (MidV and AIP).

There are several reasons why a registration may appear unreliable or shifts may have a wide range. The biggest reason is that different individuals have different visual and logical interpretations of the same data. Depending on that observer's training and understanding of the images used the results can and will vary. Specific image registration training can minimize these variations, but it does not eliminate them. This experiment used a variety of individuals encompassing physicists and physicians with different levels of experience in order to cast a wide net. In future investigation, radiation therapists should be included in the study design since they are responsible for the initial patient shifts and alignment at the treatment machine. For time commitment limitations, radiation therapists were unable to be included in this experiment.

Matching a bright, sharp image to a lesser quality, motion blurred image by using the MidV and CBCT or matching a motion-blurred image to a lesser quality, motion-blurred image by using the AIP and CBCT, results in high subjectivity in the image alignment. However, the comparative range of interpretation is only clinically relevant if it exceeds the PTV margin (if no changes in breathing are assumed).

Registrations displaying poor intra-observer and inter-observer reproducibility might indicate that the images lack obvious and clear landmarks for alignment. However, if only one observer's registrations have relatively large intra-observer variability or are outliers to other inter-observer registrations, then that observer's reliability may be compromised. It is crucial that proper training is given to those responsible for patient alignment that allows for clear, consistent practice. When an adequate reference image is selected, staff is trained properly, and the PTV margin expansion is designed correctly [4][5], individual alignments should not exceed the PTV margin.

Population Results

Over the population, no significant differences between the MidV and AIP registrations were found. The standard deviations for both MidV and AIP data sets (intra-observer and inter-observer) were, on average over all 16 patients, relatively consistent and small (≤ 2 mm). This would indicate that, over this small population, there is no clinically relevant difference between using the MidV reference image or the AIP reference image for lung cancer patient alignment.

However, specific patients were found to have relatively large differences in shifts when using the MidV reference image or the AIP reference image. Since accuracy of alignment is not a quantifiable indicator of quality through this experiment, a proper agreement between the MidV/CBCT and AIP/CBCT registrations relative to each other is all that can be compared. Additionally, some patients had relatively improved reproducibility using one reference image over the other. For these reasons, patients and their fractions were examined individually.

Per Patient, Per Fraction Results

MidV and AIP Shift Differences

The data in Figures 2 and 3 indicate which patients had extreme differences between the AIP and MidV registrations (black lines), the extent to which the average MidV registration over all observers and the average AIP registration over all observers agree (blue lines), and how consistent each MidV and AIP registrations were per observer, averaged over all observers included in the study (green lines). All three lines are important because the data must speak to an expected range as well as reproducibility/consistency. Consistent registration differences between the two references could indicate a systematic shift caused by the selected MidV reference image not being an accurate surrogate of the mid-ventilation position. Shift differences also may be due to the inherent visual differences between the MidV and AIP image. This information shows the potential for clinically relevant differences in alignment when using either method.

The majority of patients had similar shifts when using the AIP or the MidV reference image which indicates little clinically relevant consequences when using one reference image over the other in terms the location of alignment. A few patients had more significant differences between the two reference images. This paper cannot discuss accuracy; however, the difference

between registrations from different reference images should not exceed the margin range. This would indicate that one or both reference images could result in a misalignment. No difference in magnitude, maximum, or average, exceeded 1cm; therefore, there is no clinically significant difference when using ≥ 5 mm margins. If reduction of margins is considered, a potential for clinically significant difference with the use of 3 mm margins was found for a few patients (Patients 1, 8, and 11). All three patients had tumor motion of 1.4 cm and higher. However, other patients with tumor motion above 1.4 cm did not have large differences between the AIP and MidV registrations, so tumor motion alone does not appear to be a reliable indicator for differences between the two reference images. No other examined physical characteristics were shared across the three patients, and no other correlations or potential causes for the large differences were found. Regardless of any differences, no recommendations for utilizing either reference image can be made on this information. Since the cause for the difference in alignments cannot be strongly linked to physical characteristics, visual differences, or a systematic shift it is recommended that 3 mm margins not be used until more conclusive data can be collected

Inter-Observer Variability Comparison

Figure 4 shows how much more the shifts of one reference image set varied over the other. If the difference per fraction was approximately zero, then the two methods has similar variation. If the difference per fraction was higher than zero, the MidV method resulted in higher inter-observer ranges by that amount. If the difference per fraction was lower than zero, the AIP method resulted in higher inter-observer ranges by that amount. This easily shows if a patient benefited from using one method over the other in terms of inter-observer variability. A few patients had relatively lower variability when one reference image was used over the other. In order to determine if a reference image consistently has lower variability than the other when common issues are present, correlations between variability and the reported physical characteristics were made.

Correlations Between Registration Variability and Common Issues

In Figure 5, both MidV and AIP registrations display a moderate correlation (R= 0.669 and R= 0.633, respectively) between inter-observer differences and the patient's setup rotation angle. The MidV registrations correlate linearly slightly better than the AIP registrations, but

both show a general positive trend. From these plots, it is seen that not all patients with large differences between inter-observer registrations were rotated during their worst fraction, but it is seen that all patients that were rotated greater than 4° during their worst fraction exhibited large differences (minor exception of patient in AIP set whose maximum was just below 6 mm). For this reason, rotation should be monitored and corrected, especially if using margins ≤ 3 mm. This data also might show that AIP reference images may slightly have more immunity to registration difficulties when rotations are present. This can be attributed to the motion-blurred nature of both the CBCT and AIP images. These leave fewer specific features and information for focused visual match, so instead the general area is aligned. A larger patient cohort is needed to have any statistical significance of this claim.

Figure 6 shows the linear correlation is seen to be smaller for the MidV data set and both show a positive trend implying that registration variability does increase with tumor motion. Correlations using the average inter-observer variability show comparable results. In general, although the AIP data linearly correlates more, it is shown that tumor motion affects the inter-observer variability approximately the same for either reference image. The patient with the highest motion (1.8 cm, Patient 8) had more consistent results when using the MidV reference images. The high tumor motion resulted in a highly blurred AIP reference image which likely led to the larger registration variability. Patients with motion 1.8 cm or greater may benefit from using a MidV reference image over the AIP. More patients are needed to make any significant claim.

Finally, no correlation between tumor volume and registration variability was found. No plots correlating tumor volume and registration variability are shown. It was observed that all patients with tumor volumes > 80 cc had at least one fraction with relatively high registration variability using either registration method. Tumors with larger volumes gave the observer more points of interest to choose from when aligning; therefore, the subjective nature of alignments is increased. Additionally, a patient with a small tumor (2.1 cc, Patient 8) with 1.8 cm of motion was shown to have larger variability when using the AIP reference image. More patients are required to determine if only small tumors with high motion have less registration variability when using the MidV reference image.

Significance and Clinical Relevance

A single patient's data indicated that the MidV reference image may result in less registration variability for patients with tumors whose motion is ≥ 1.8 cm. This implies that the MidV reference image may be more clinically appropriate for patients with high tumor motion. More patients would be needed to make a significant claim.

For the majority of patients, the AIP reference image had slightly less registration variability when those patients were rotated. This implies that the AIP reference image may facilitate more reproducible registration when patients are rotated. More data is needed to make a significant claim.

A few patients had relatively large differences between registration using either method; these results indicate that there is the potential for clinically significant alignment differences when small margins are used. The assumption that the MidV reference image and the AIP are representative of each other is not always seen to be accurate in terms of clinical significance. Guided breathing may allow for better agreement and a more reliable assumption.

Previous Works

The results of this study are consistent with previous findings. Komomae et al. [20] concluded that the inter-observer variability using the AIP could be significant (1.43 mm difference in SI direction using a phantom study); the MidV was not analyzed. Oechsner et al. [21] concluded that there could be a clinically significant difference between the AIP and the MidV reference images, and tumor motion was correlated with shift differences. These findings further indicate that there may be clinically relevant shift differences when using the MidV or AIP reference image for high motion tumors. Additionally, Oechsner et al. found that the average inter-observer variability when using the AIP reference image was the smallest and the maximum inter-observer variability was highest for the MidV reference image. These results varied from this paper's findings as both reference images had comparable average variability and maximum variability. This could be due to patient specific tumor differences (tumor size was not examined in the study) or differences in observer subjectivity.

Future Work

Since reference images are also used in planning, their dosimetric impact should be investigated as well. The sample size in this paper is too small to make significant statistical claims and correlations; more patient data should be added to the cohort. In future iterations, radiation therapists should be included as observers.

Guided breathing may be able to help reduce these uncertainties and variabilities, but more research and a practical protocol must be tested before implementation at VCU. The uncertainties in this experiment would greatly diminish if the same processes were repeated for patients who underwent guided breathing. The experiment would more easily be able to determine which reference image has the least amount of intra-observer and intra-observer variably while being able to strongly rely on the trustworthiness of the reference images' definitions. Ultimately, the determination of whether one reference image is better than the other would be dependent on the visual differences of the two reference images and the physical characteristics seen in Table 2. If proper guided breathing protocols are established, this experiment can be performed again and compared. Until then, either reference image would result in clinically appropriate treatment if 5 mm margins are used.

Conclusion

Overall, no clinically relevant differences were discovered between using the MidV method and AIP method if ≥ 5 mm margins are used. If the decision to reduce margins further is made, the difference between the two reference images may be more clinically relevant.

If margins are reduced to 3 mm, for example, the clinic must be able to align the patients with more fidelity and less rotation, as rotations led to ≥ 6 mm discrepancies between registrations of the same type and between the two methods. The AIP reference image showed a marginally increased reliability when patients were rotated, but not enough to compensate for all cases. For tumors with high motion (≥1.8 cm), the MidV reference image seems to perform better over the AIP based on the differences in registration reproducibility from the limited patient cohort included in this study. However, differences in reproducibility were small; therefore, the added complexity to the workflow of having different reference images for different patients is not justified based on this data.

Furthermore, guided breathing would be suggested to improve consistency in respiration between the simulation scan and pre-treatment CBCTs. This would make the breathing captured by the simulation scan, regardless of the reference image selected, more representative of the breathing at treatment which would improve the alignment reliability.

Appendix

Intra-Observer Results

Table A 1. Table showing magnitude of the average intra-observer standard deviation in all three directions (LAT, AP, and SI) and the magnitude of the maximum intra-observer standard deviation in all three directions per patient for all three intra-observer trials.

	Ave MidV Intra StDev (Max StDev) (mm)	Ave AIP Intra StDev (Max StDev) (mm)	Ave exRef Intra StDev (Max StDev) (mm)
1	1.6 (1.8)	1.8 (2.2)	0.7 (0.8)
2	1.2 (1.9)	0.7 (0.8)	0.8 (0.9)
3	1.3 (2.0)	1.3 (2.1)	1.1 (1.4)
4	0.9 (1.2)	0.8 (1.0)	1.4 (1.7)
5	0.8 (1.1)	0.7 (0.8)	0.9 (1.1)
6	2.3 (3.0)	0.7 (0.8)	1.5 (1.6)
7	2.4 (2.9)	1.7 (2.0)	1.1 (1.7)
8	1.0 (1.8)	1.3 (1.4)	1.3 (1.5)
9	0.2 (0.3)	0.2 (0.3)	0.4 (0.5)
10	0.4 (0.5)	0.4 (0.4)	0.6 (0.9)
11	2.0 (2.2)	1.3 (1.8)	1.4 (1.5)
12	0.5 (0.7)	0.8 (0.9)	1.0 (1.3)
13	1.3 (2.1)	0.8 (1.3)	1.0 (1.3)
14	0.9 (1.3)	1.2 (1.8)	1.1 (1.1)
15	1.5 (2.4)	0.8 (1.1)	1.5 (1.7)
16	1.3 (1.7)	0.9 (1.1)	1.1 (1.4)

MidV and AIP Ranges

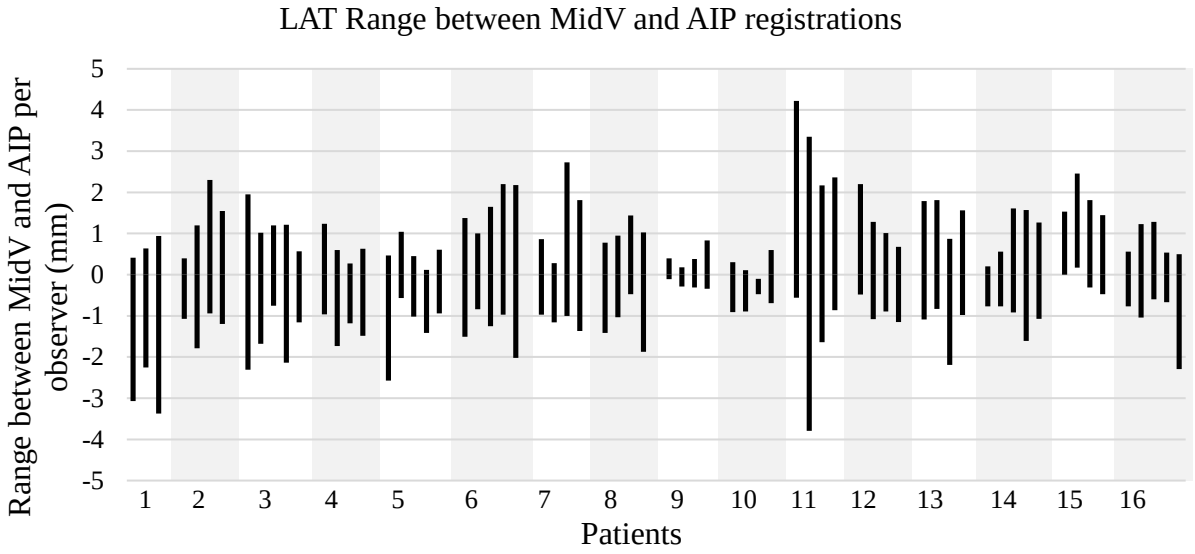

Figure A 1. Range plot of the difference between lateral shifts for the MidV and AIP registrations. Upper bound indicates what value an individual's LAT MidV shift was highest over their LAT AIP shift. Lower bound indicates what value an individua's LAT AIP was highest over their LAT MidV shift.

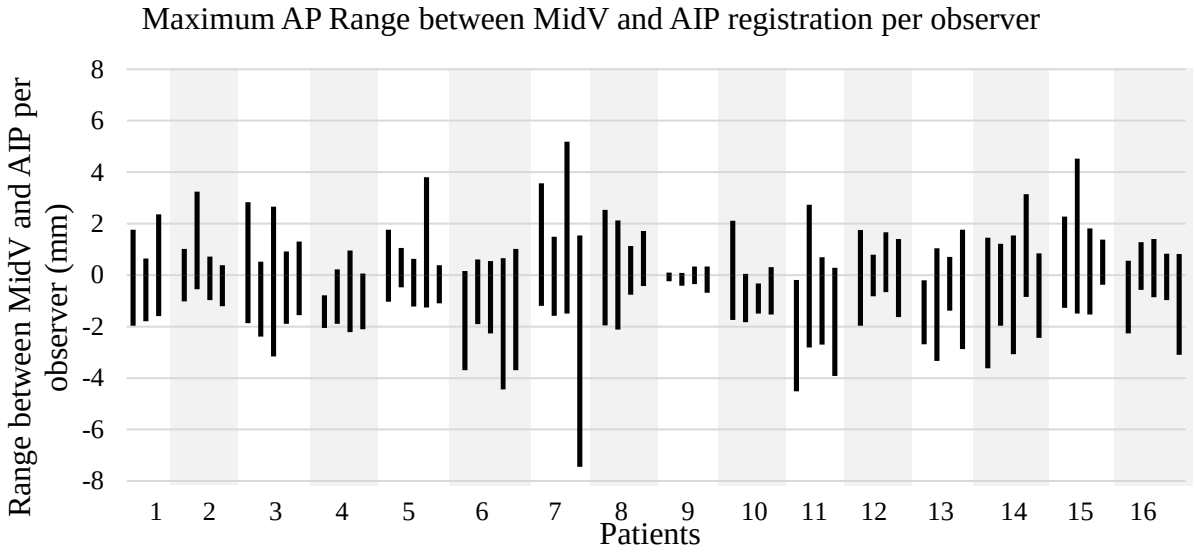

Figure A 2. Range plot of the difference between anterior-posterior shifts for the MidV and AIP registrations. Upper bound indicates what value an individual's AP MidV shift was highest over their AP AIP shift. Lower bound indicates what value an individua's AP AIP was highest over their AP MidV shift.

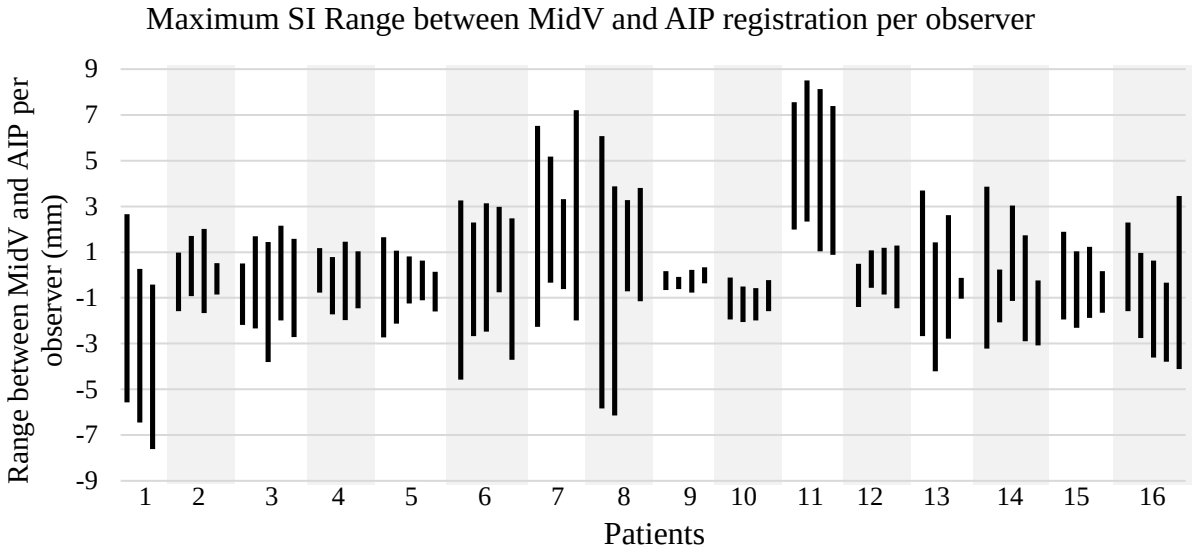

Figure A 3. Range plot of the difference between superior-inferior shifts for the MidV and AIP registrations. Upper bound indicates what value an individual's SI MidV shift was highest over their SI AIP shift. Lower bound indicates what value an individua's SI AIP was highest over their SI MidV shift.

Inter-Observer Registration Variability

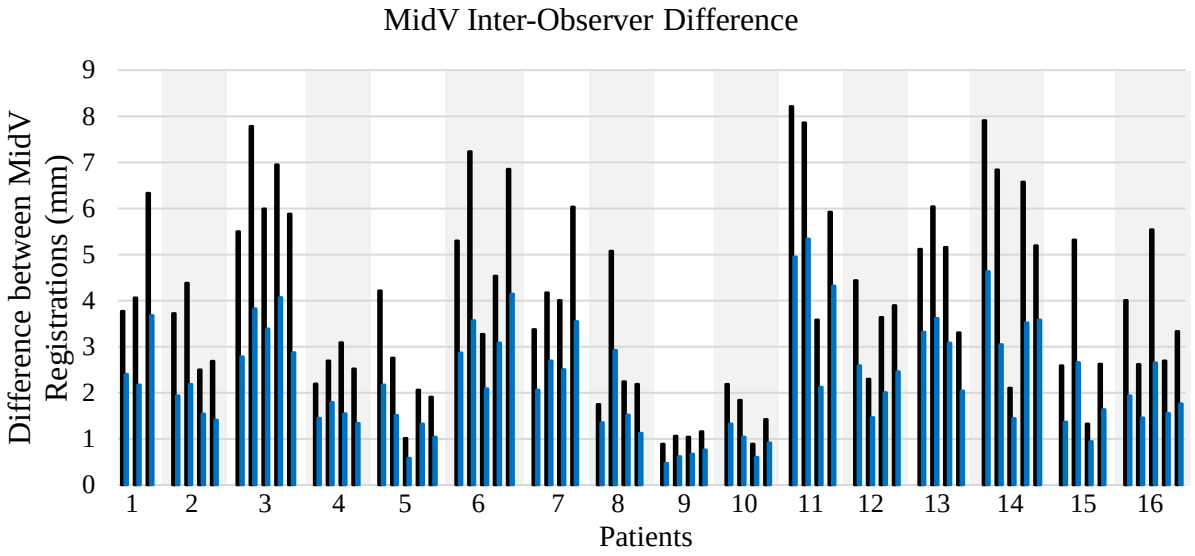

Figure A 4. Graph of the maximum difference between inter-observer registrations using the MidV reference image (Eq. 3, black lines) and the average difference between inter-observer registrations using the MidV reference image (blue lines) for all fractions per patient. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background. Patients 1, 3, 6, 7, 11, 13 and 14 have at least a single fraction with inter-observer variability above 6 mm.

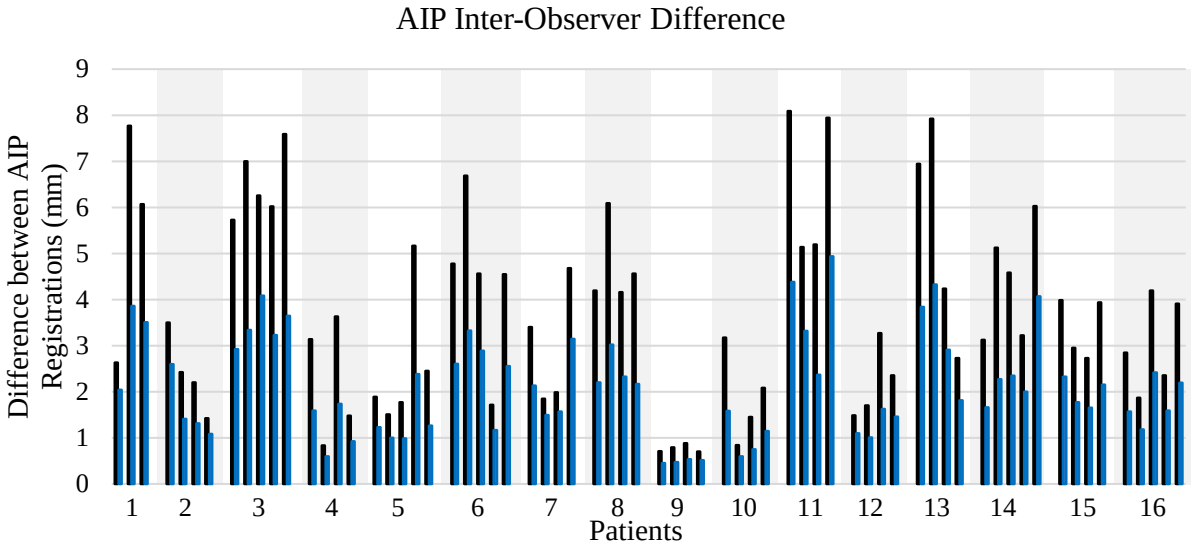

Figure A 5. Graph of the maximum difference between inter-observer registrations using the AIP reference image (Eq. 3, black lines) and the average difference between inter-observer registrations using the MidV reference image (blue lines) for all fractions per patient. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background. Patients 1, 3, 6, 7, 11, 13 and 14 have at least a single fraction with inter-observer variability above 6 mm.

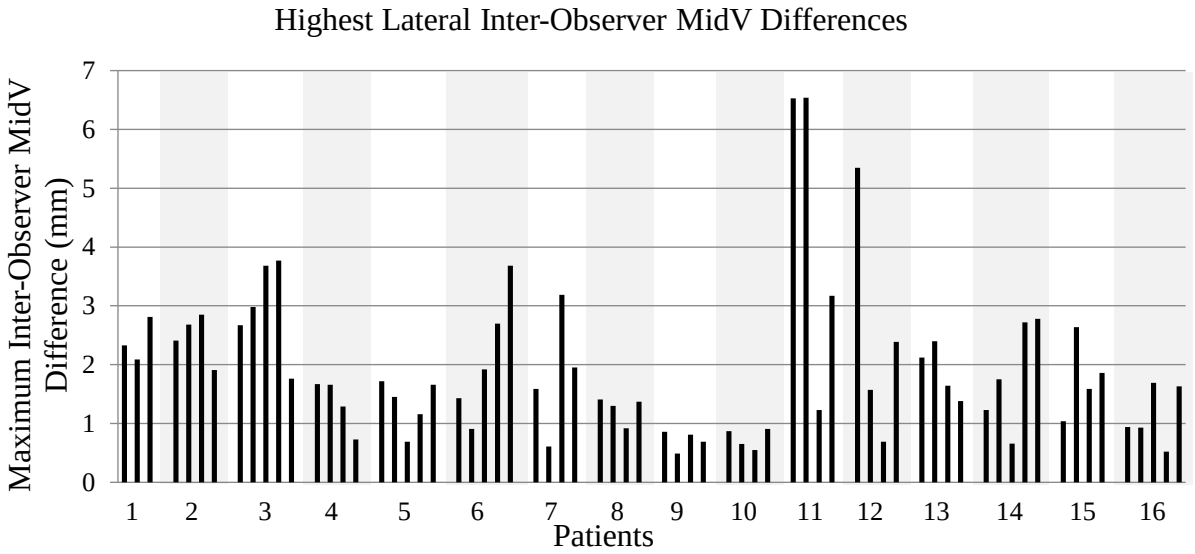

Figure A 6. Figure shows the highest inter-observer difference in the lateral direction in the MidV data sets for all patients. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background.

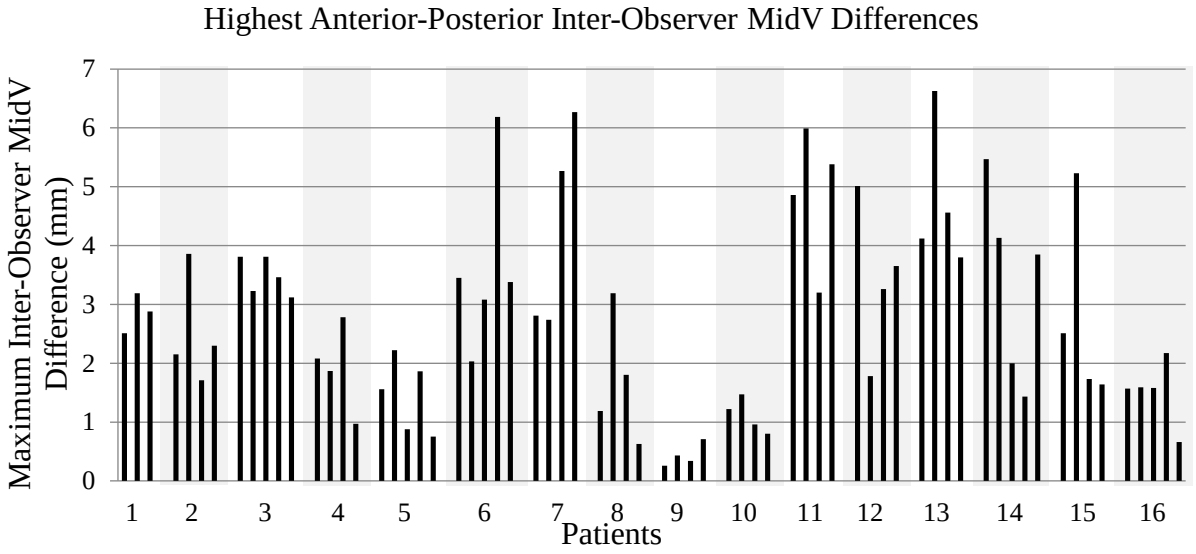

Figure A 7. Figure shows the highest inter-observer difference in the anterior-posterior direction in the MidV data sets for all patients. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background.

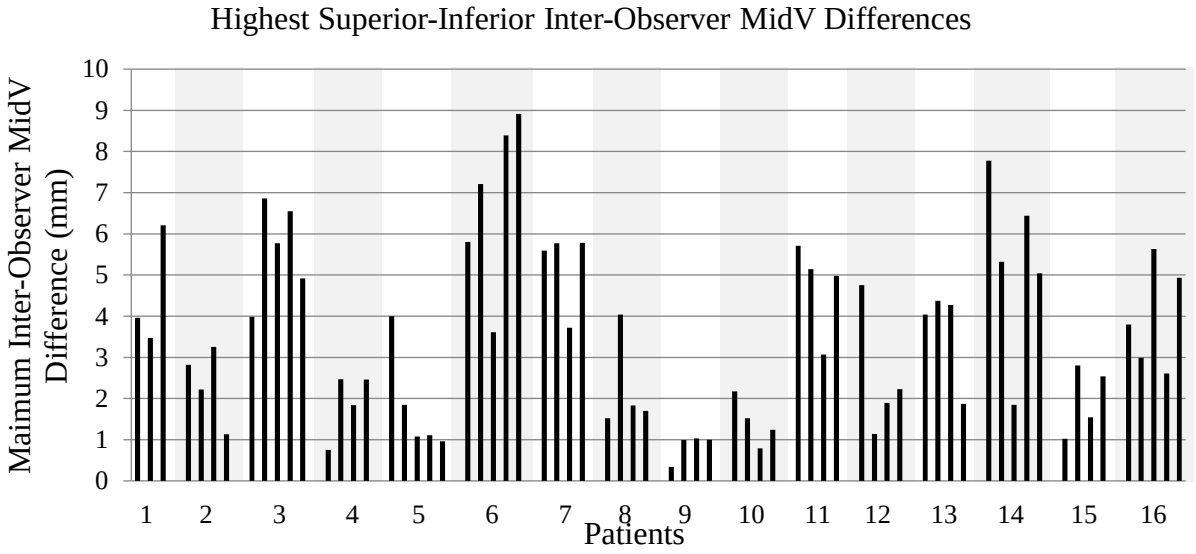

Figure A 8. Figure shows the highest inter-observer difference in the superior-inferior direction in the MidV data sets for all patients. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background.

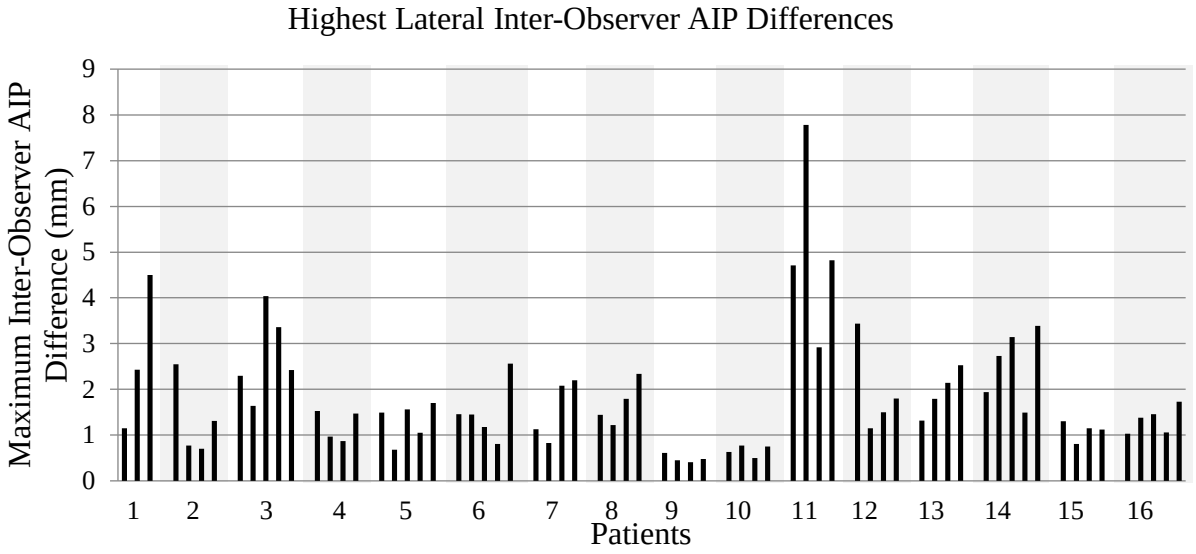

Figure A 9. Figure shows the highest inter-observer difference in the lateral direction in the AIP data sets for all patients. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background.

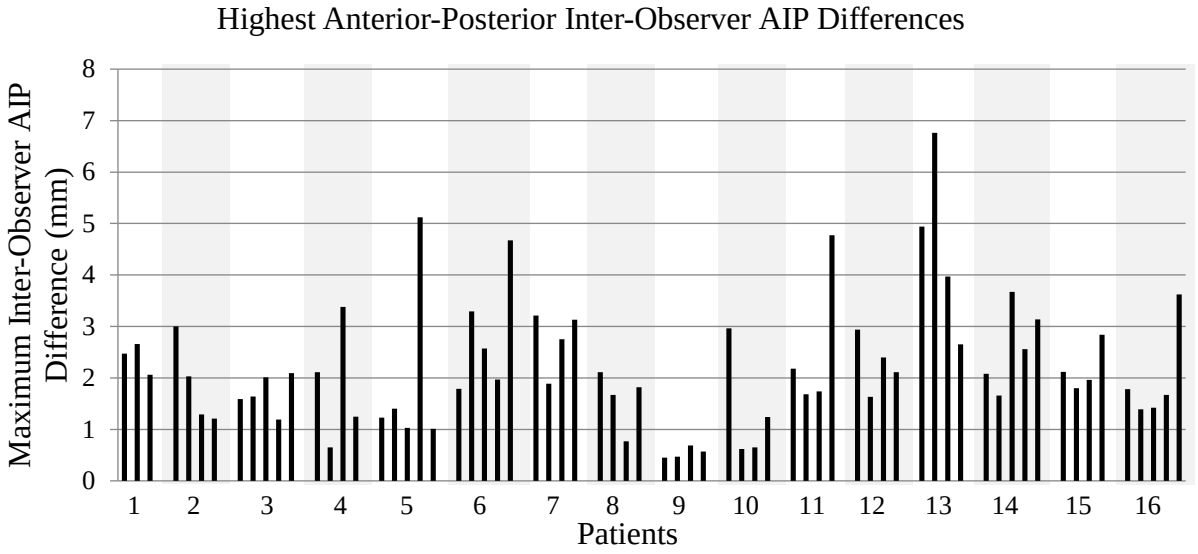

Figure A 10. Figure shows the highest inter-observer difference in the anterior-posterior direction in the AIP data sets for all patients. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background.

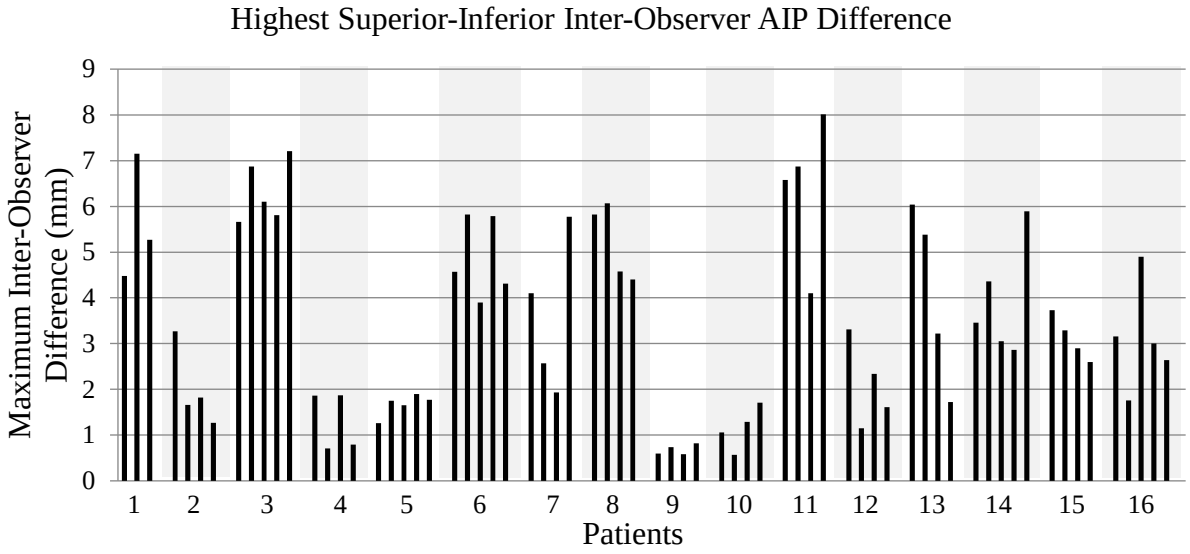

Figure A 11. Figure shows the highest inter-observer difference in the superior-inferior direction in the AIP data sets for all patients. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background.

Interquartile Ranges

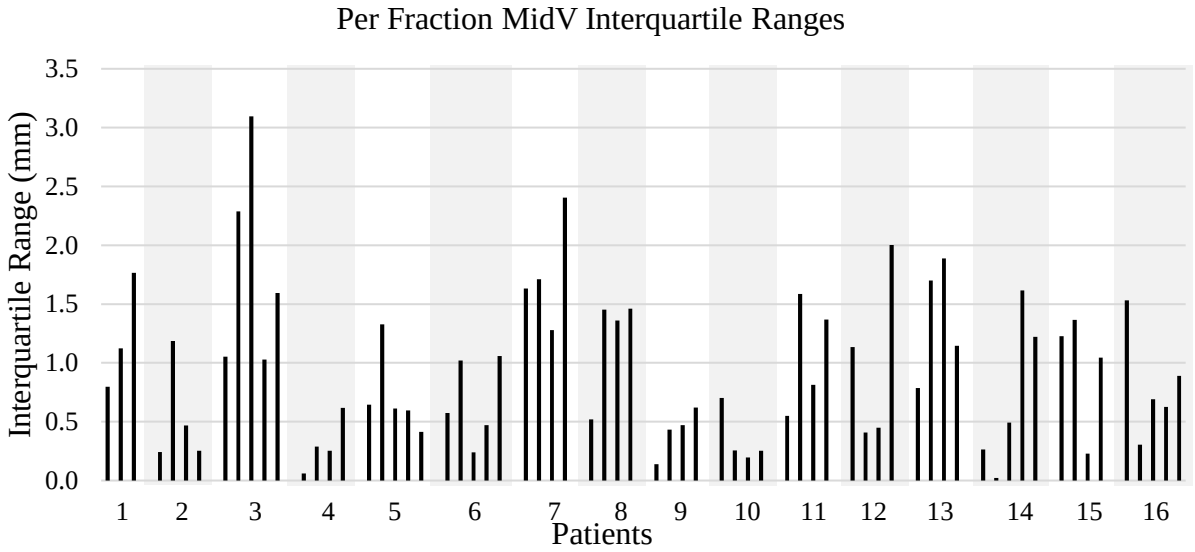

Figure A 12. Figure shows the MidV interquartile range per fraction for all patients. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background.

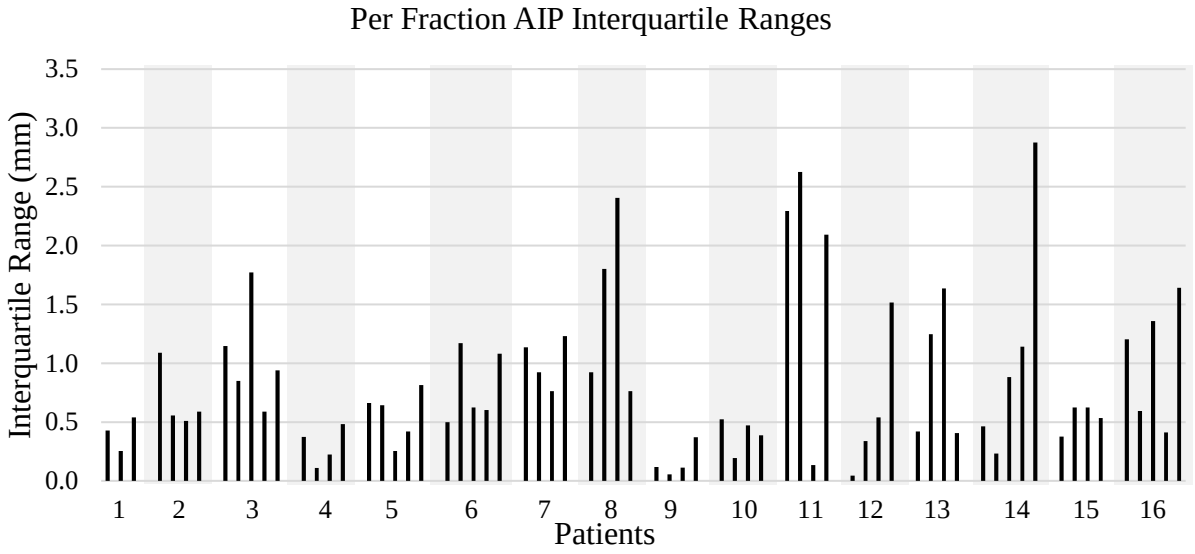

Figure A 13. Figure shows the AIP interquartile range per fraction for all patients. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background.

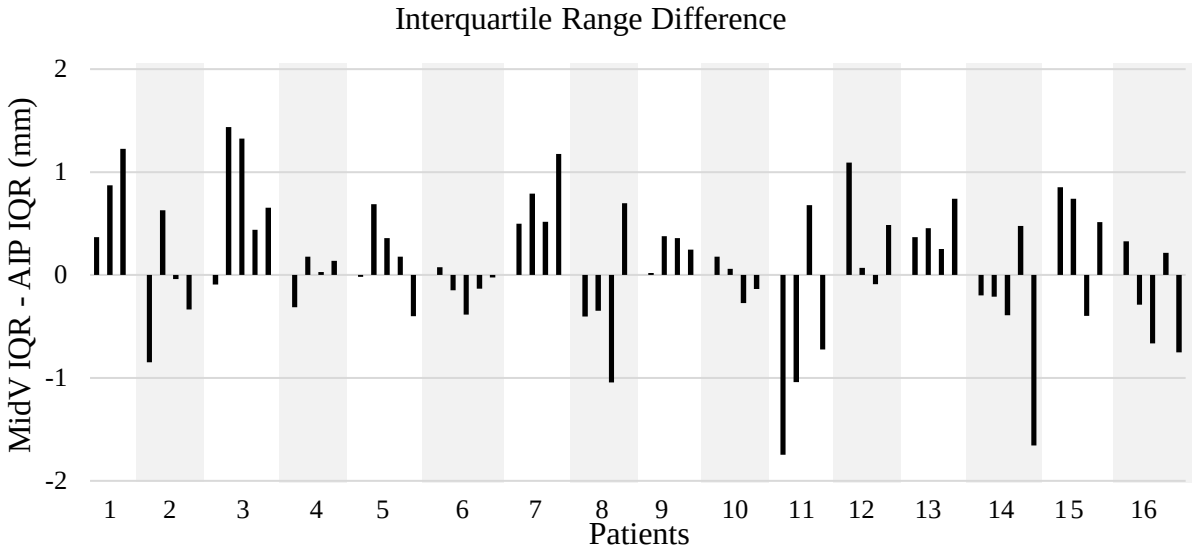

Figure A 14. Graph of the difference in Interquartile ranges (MidV-AIP). A positive number indicates the IQR is larger for the MidV fraction. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background.

Additional Correlations

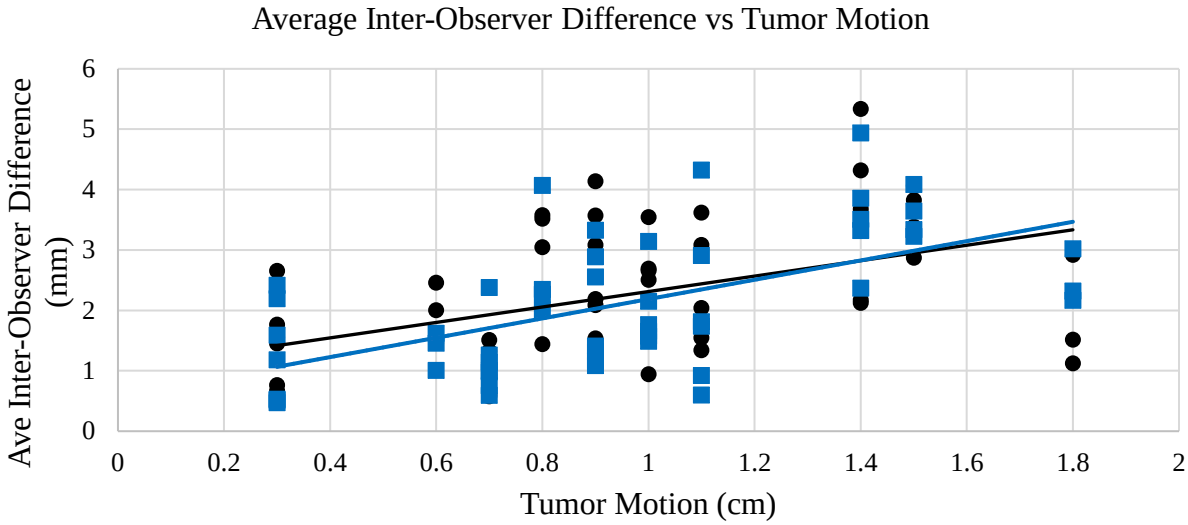

Figure A 15. Scatter plot of magnitude of the estimated tumor motion (Table 2) vs the average MidV (black circles) or AIP (blue squares) inter-observer variability (Eq. 3) for the all fractions. Can indicate if/how inter-obs. variability can be affected by tumor motion. Shows positive trend for both reference images and comparable correlations. All points with a variability > 6 mm have tumor motion > 0.8 mm. R values = 0.453 and 0.578 for the MidV and AIP data sets, respectively.

Significant Registration Differences

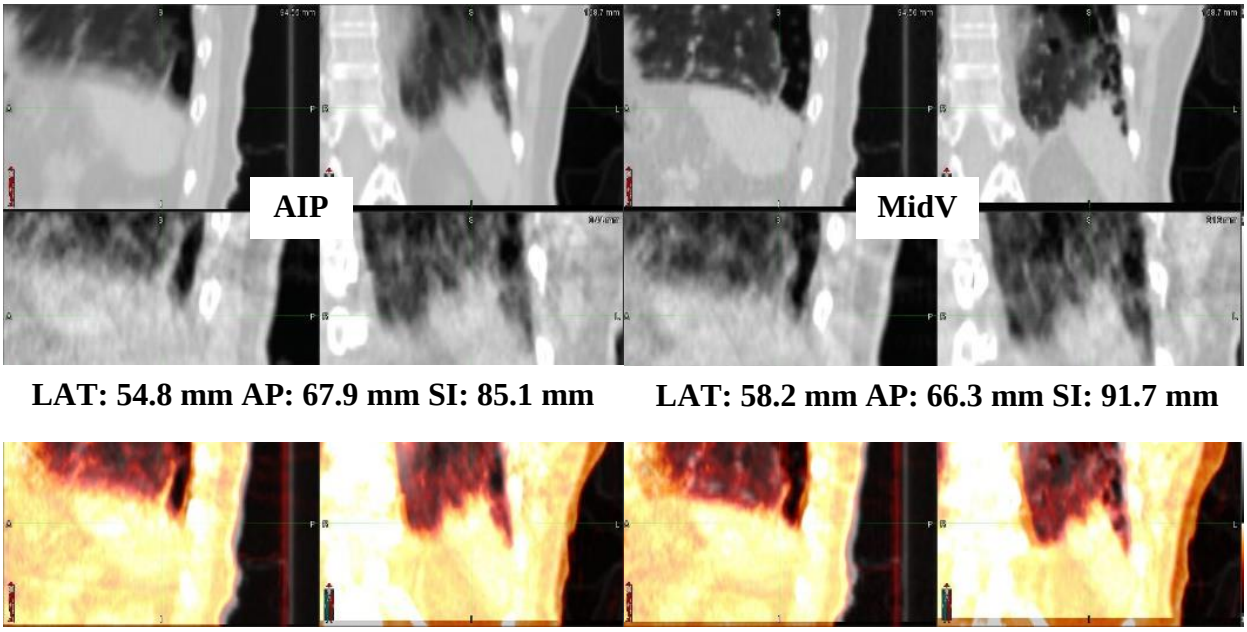

Figure A 166. Visual demonstration of a single observer's registrations (AIP/CBCT vs MidV/CBCT) exceeding a 6 mm difference. Both alignments are reasonable, yet their shifts vary outside a 3 mm margin.

Patient Specific Discussion

In this section, each patient will be examined closely and individually. Since the patient cohort was small to determine many statically relevant results, this paper elects to examine the specific errors and traits that each patient had.

Patient 1

Patient 1 had a 90.6 cc tumor and did not show signs of rotations for the two worst fractions. The tumor had 1.4 cm of motion and no artifacts were present in the worst two fractions. All three fractions had at least one observer register the MidV and AIP differently by 6 mm or more. This indicates that all three fractions had the potential for a quality registration failure by using one or both reference images. Inter-observer variability exceeded 6 mm for the 3rd fraction for both methods and exceeded it in the 2nd fraction by using the AIP method. The MidV followed more closely with the exRef alignment for all three fractions by ~1.7 mm on average.

Based on the above data, the cause of the quality failures was the high motion of a large tumor and blurred nature of the AIP image. Both methods showed potential for failure if 3 mm margins had been used. Like all patients in this cohort, 5 mm margins would be clinically sufficient. Based off of registration variability, the MidV reference image may have best suited this patient, but the fidelity of its assumptions is not inherently reliable due to the differences seen between the MidV and AIP registrations.

Patient 2

Patient 2 had a 5.3 cc tumor and virtually did not show the presence of rotation for the two worst fractions. The tumor had 0.9 cm of motion and no artifacts were noticed. No significant difference between any two relevant/comparable variables were found. Either reference image is recommended.

Patient 3

Patient 3 had a 45.5 cc tumor and showed the presence of rotation during their worst two fractions (5° for both). The tumor had 1.5 cm of motion and no artifacts were noticed. No MidV/AIP comparison resulted in a difference above 6 mm. With the exception of the first fraction, all fractions had significant inter-observer variability when using at least one of the data sets. For the fractions that only showed one method > 6 mm, the other method still showed a

maximum of ~5.8 mm. No significant differences were found when comparing the exRef alignment.

Based on the above data, rotation may have had a clear impact on the two worst fractions. It may have been present for all fraction. Rotation is likely the cause of the failure of the MidV method. The high tumor motion may have led to the failure of the AIP method. Since results show the potential for both methods to fail, 3 mm margins would not be recommended for this patient unless rotation could be corrected, and motion could be managed. Interestingly, no observer registered the two reference images by exceeding a 6 mm difference. This implies consistency using either method amongst the same observer and implied a good relationship between the AIP and MidV.

Patient 4

Patient 4 had a 9.5 cc tumor and virtually did not show the presence of rotation for the two worst fractions. The tumor had 1.1 cm of motion and no artifacts were noticed. No significant difference between any two relevant/comparable variables were found. Either reference image is recommended.

Patient 5

Patient 5 had a 5.6 cc tumor and showed virtually no rotation during either investigated fraction. The tumor had 0.7 cm of motion and no artifacts were noticed. No significant difference between any two relevant/comparable variables were found. Either reference image is recommended.

Patient 6

Patient 6 had a 15.5 cc tumor and showed rotation of 8° during its worst MidV and AIP fraction (fraction 2). The tumor had 0.9 cm of motion and artifacts were present during the worst fraction. No MidV/AIP comparison resulted in a difference above 6 mm. Fraction 2 and 5 had excessive inter-observer MidV variability; fraction 2 had excessive inter-observer AIP variability as well. Both reference images followed the exRef alignment closely.

Based on the above data and the fact that not all fractions resulted in large differences, it is implied that a situational error like rotation was the root cause for fraction 2. Rotation again may have led to the wide range of MidV registrations in fraction 5, but it was not recorded. The AIP appears to have been able to compensate in fraction 5, but not for fraction 2. Furthermore, the presence of artifacts could have led to uncertainties when aligning. Rotation is a potentially

correctable issue; unfortunately, artifacts are less predictable. Rotational corrections should be mandatory when using 3 mm margins.

Patient 7

Patient 7 had a 22.1 cc tumor and had a rotation of 4° for the fourth fraction. The tumor had 1 cm of motion and artifacts were present during the fourth fraction. No MidV/AIP comparison resulted in a difference above 6 mm. Only the fourth fraction had inter-observer variability over 6 mm using the MidV reference image. No > 6mm issues were found using the AIP reference image. Both the MidV and AIP registrations followed the exRef similarly.

Based on the data above, rotation is likely the cause of the failure when using the MidV for the fourth fraction. The AIP data reflected a more consistent alignment from the group, and therefore, its variability was not impacted by the rotation. Again, if 3 mm margins are to be used, rotation must be corrected for. No other outstanding physical characteristics are present to cause this failure. This also weakly implies a slight advantage that using the AIP might have when patients are rotated.

Patient 8

Patient 8 had a 2.1 cc tumor and showed virtually no rotation during the second fraction. The tumor had 1.8 cm of motion and artifacts were present during the investigated fractions. Fractions 1 and 2 had a MidV/AIP difference exceeding 6 mm but had less issues for the 3rd and 4th fractions. This implies a fair relationship between the MidV and AIP reference image in terms of definition. The majority of observers had > 6 mm differences for both fractions. Fraction 2 had inter-observer variability greater than 6 mm when the AIP reference image was used. No significant issues were seen for the MidV data set, but it's highest was for fraction 2. The 1st fraction had an exRef and AIP difference over 5 mm; indicating an outlier and issue.

Based on the data above, this patient had specific issues during the first two fractions. The MidV reference image was seen to perform appropriately and better for all fractions. The data suggests that the high tumor motion and small volume of the tumor were to blame for the failure of the AIP. The difference between the exRef and AIP in the first fraction implies that the AIP and CBCT did not visually correlate well for the first two treatments. Potentially, the breathing may have been erratic during simulation or during treatment. The average tumor location may have changed with respect to amplitude. This combined with the AIP's visual uncertainty could cause such shift. The MidV registrations may have not suffered from this due

to having a clear “snapshot” image which a group of individuals could more easily register consistently. It is likely the breathing returned to normal for the second two fractions. This patient is a good example of why guided breathing should be performed when using either reference image. Since the tumor was small and had a high magnitude of motion, it makes sense that there would not be a lot of visual information in the AIP to use, and therefore, it is important that the treatment day be as consistent as simulation. Overall, when dealing with small, high motions tumors, guided breathing may be an appropriate decision even if using the MidV reference image.

Patient 9

Patient 9 had a 1 cc tumor and showed virtually no rotation for the worst fraction. The tumor had 0.3 cm of motion and no artifacts were noticed. No significant difference between any two relevant/comparable variables were found. Either reference image is recommended.

Patient 10

Patient 10 had a 2.3 cc tumor and showed virtually no rotation for the worst fraction. The tumor had 0.7 cm of motion and no artifacts were noticed. No significant difference between any two relevant/comparable variables were found. Either reference image is recommended.

Patient 11

Patient 11 had a 26.1 cc tumor and showed 10° of rotation for the worst fraction. The tumor had 1.4 cm of motion and showed the presence of artifacts in the first fraction. Most observers registered the MidV and AIP reference images over 6 mm different for all four fractions. The MidV inter-observer variability exceeded 6 mm for fractions 1 and 2 (5.8 mm for fraction 4), and the AIP inter-observer variability exceeded 6 mm for fractions 1 and 4. The interquartile ranges were much larger for the AIP data set (with the exception of fraction 3). This implies the AIP shifts were more spread out. The average MidV registrations showed a much larger difference from the exRef.

Based on the information above, rotation likely caused some of the large variability seen in fraction one. Rotation could have been present for all four fractions. The relatively large difference between the MidV and exRef implies that the MidV was not a good representation of the average tumor location during Sim. This accounts for some of the discrepancies between the two reference image shifts. The presence of artifacts and large tumor motion likely played a role in the variability using both images. The conclusion that 3 mm margins should not be used until

rotation can be corrected, until guided breathing's effects are investigated, or in the presence of artifacts remains.

Patient 12

Patient 12 had a 7.0 cc tumor and virtually did not show the presence of rotation for the two worst fractions. The tumor had 0.6 cm of motion and no artifacts were noticed. No significant difference between any two relevant/comparable variables were found. Either reference image is recommended.

Patient 13

Patient 13 had a 166.5 cc tumor and showed 4° of rotation during the worst fraction. The tumor had 1.1 cm of motion and no artifacts were noticed. All MidV/AIP registrations agreed under a 6 mm difference. The MidV inter-observer variability exceeded 6 mm for the 2nd fraction. The AIP inter-observer variability exceeded 6 mm for the 1st and 2nd fractions. The IQ ranges implied that the AIP data was more evenly spread. Both reference average registrations agreed well with the exRef.

Based on the above information, rotation again may have led to the variability in the 2nd fraction. It is likely that the large size amplified the relatively smaller rotation angle compared to other rotated fractions. The information implies that large tumors, regardless of tumor motion, may not be suitable for 3 mm margins if rotation cannot be corrected.

Patient 14

Patient 14 had a 111.7 cc tumor and showed virtually no rotation during the worst two fractions. The tumor had 0.8 cm of motion and no artifacts were noticed. Comparison between MidV and AIP registrations did not show difference above 6 mm. MidV inter-observer variability exceeded 6 mm for 1st, 2nd, and 4th fractions. AIP inter-observer variability was 6 mm for the 5th fraction. Both reference average registrations agreed well with the exRef.

Base on the data above, the only outstanding characteristic (that was investigated) which could account for the large inter-observer discrepancies could be the size. It is also noted that the tumor shape might have aided in this variability. See Figure A 17. The AIP performed much better than the MidV data set in this patient's case, although it still had an issue with the 5th fraction. Based off the data and Fig. A 17, the clear edges of the MidV may have caused its issued. Observers may have chosen different points about the large tumor to align to. Although,

the AIP is not significantly more blurred than the AIP due to the small tumor motion. It can be weakly concluded that large tumors may not benefit from the use of 3 mm margins.

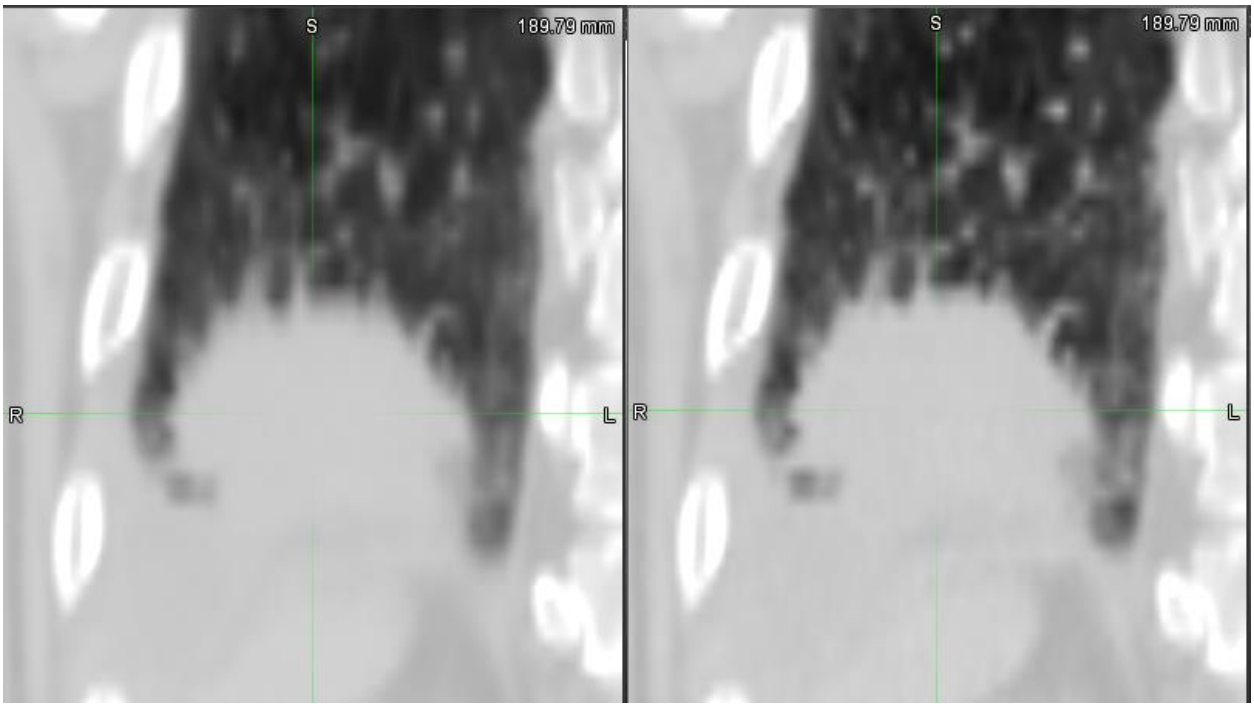

Figure A 177. Patient 14's tumor. Left image shows AIP of the tumor; right image shows the MidV of the tumor. The "spikey" anatomical nature of the tumor may account for registration variability in addition it its size.

Patient 15

Patient 15 had a 33.6 cc tumor and virtually did not show the presence of rotation for the two worst fractions. The tumor had 1.0 cm of motion and no artifacts were noticed. No significant difference between any two relevant/comparable variables were found. Either reference image is recommended.

Patient 16

Patient 16 had a 50.1 cc tumor and show 3° rotation for the worst fraction. The tumor had 0.3 cm of motion and no artifacts were noticed. No significant difference between any two relevant/comparable variables were found. Either reference image is recommended.

Exhale Reference Alignment

To compare breathing cycle variability and its effect on the two methods, a unique system of comparisons was created. The main idea of this method was to yield a consistent and reproducible reference alignment that could be compared to the registrations between the AIP and CBCT or between the MidV and CBCT of each fraction. The difference between the

alignments would indicate how much the breathing cycle varied between simulation (Sim) and treatment.

To achieve a reproducible reference, the most exhale position (superior edge) of the tumor in the Window/Level adjusted CBCT (called the Pseudo-MIP or PMIP) was used for alignment with the most exhale position (superior edge) of the tumor in the MIP. Therefore, the reference alignment is referred to as the Exhale Reference alignment (exRef). It has been shown in literature that the most consistent position of a tumor throughout all breathing cycles is the exhale position [7][8]; therefore, it was chosen as the stable comparison point.

exRef Method for Quantifying Breathing Cycle Variation

An "exhale reference" alignment (exRef) was also determined by registering the maximum intensity projection (MIP) with pseudo-MIP (PMIP) of each CBCT (MIP/PMIP) at the superior edge of each (exhale position). The pseudo-MIP was created by adjusting the window/level of the CBCT to brightly show the tumor throughout its various respiration-dependent locations during the acquisition. In essence, the contrast would be set to show brightly everywhere the tumor was during its respiratory cycle acquired by the CBCT.

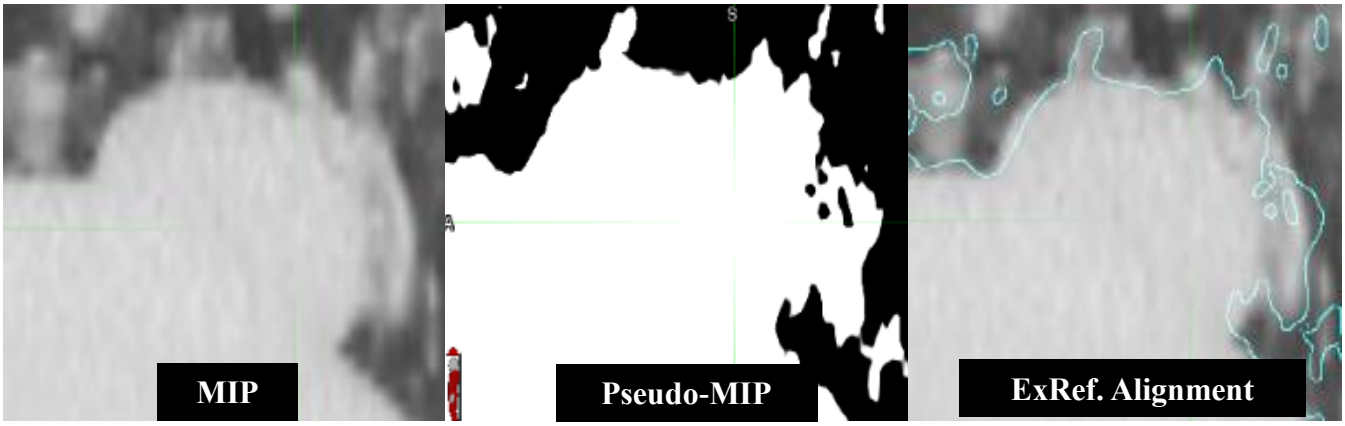

Figure A 18. Shows the maximum intensity projection (MIP), pseudo-MIP (PMIP), and the exhale reference (exRef) image registration between the two. Magnification of all three images is consistent. The PMIP is the CBCT obtained before treatment and has its Window/Level adjusted to have full brightness over the tumor's entire respiration path; this mimics the concept of the MIP image. Blue lines in the exRef image represent the superior-edge matching between the MIP and PMIP. Showed to visually demonstrate the alignment method.

Pseudo-MIP Method Justification and Limitations

The use of the MIP and the creation of the pseudo-MIP to find an exhale reference registration was decided on due to the desire to compare the AIP and MidV registrations with a more consistent, reproducible alignment. Furthermore, this comparison could potentially show

the variability of the breathing cycle between simulation and treatment. MIP images are readily available for 4DCT data sets and they effectively show the amplitude of the breathing cycle during simulation. The exhale position has been found to be the most consistent position in the breathing cycle [7][8], and for this reason it was used to align the MIP and PMIP images. Additionally, the superior edges of the lower lung tumors were the only edges available for alignment.

The motivation for using this method is that the differences between the exRef registrations and the reference image registrations will show how much the breathing cycle changed between simulation and treatment and potentially how the image characteristics affect the registration for each treatment.

The expected relationship between the exRef alignment and MidV/CBCT alignment deteriorates when the MidV is not a good representation of the most likely tumor position at Sim. Otherwise, the shift differences should theoretically be reasonably reflective of the breathing cycle variation.

However, despite the theoretical advantages to this method, there are some practical limitations. The exact window/level of each CBCT was subject to change due to the nature of the registration method and due to the CBCT quality itself. Even though MIP's eliminate the inherent visual differences of the other two reference images, it still was up to the observer to determine when the PMIP was achieved. Although the exhale position is the most reproducible [7][8], it is not always consistent; therefore, these factors should be considered when discussing the results.

Breathing Cycle Variability's Effect on Registration Variability

Figures A 19, A 20, and A 21 deal with changes in respiratory motion. Figures A 19 and A 20 quantify how differently the average exRef registration and the average inter-observer MidV/CBCT registration and average AIP/CBCT registration compared, respectively. The average location for all three intra-observer exRef registrations and the average inter-observer registrations for either reference image data set were used as a means to get the "most likely" registration for each. It is better to use a population's data since this information should relate to a population of observers.

It is important to note that shift differences between the exRef and either of the reference image alignments do not indicate a misalignment or poor alignment using the reference images. It can however indicate a gross change in the average tumor location with respect to the amplitude. The differences seen in Figures A 19 and A 20 serve no purpose except for quantifying how much the average tumor location with respect to the amplitude changed and its impact on registering using either method.

The differences in the differences calculated for Figures A 19 and A 20 (shown in Figure A 21) give more information. These differences show how much closer one method was to the exRef than the other. A near zero difference would imply that the breathing cycle change impacted both registrations similarly. It also indicates that the two average registrations were similar to each other (i.e. how different an observer's AIP could be from their MidV registration). If a fraction has one reference image registration further from the exRef, then this would indicate a more complicated issue. No single patient had a difference in both sets greater than 6 mm (indicating that respiratory motion did not change greater than a 3 mm margin). One patient did have a difference only in the MidV data set above 6 mm.

Fractions where the average MidV registration was further from the exRef than the average AIP may happen for one main reason. Their MidV reference image might not have been an appropriate reflection of the average tumor location. Therefore, it would have a given offset in registrations from the AIP. This is an unideal situation since quality treatment depends on the reference images being faithful to their definitions. If the MidV does not represent the most common tumor position throughout the breathing cycle, then it should not be used for treatment.

Fractions where the average AIP registration was further from the exRef than the average MidV may happen for one main reason. The inherent visual obstacles observed when registering using an AIP reference image led themselves to a broad interpretation and can be difficult to confirm. These issues are likely to be seen in patients with a large range of AIP registrations and potentially high tumor motion. The assumption that the MidV reference image represented the most likely tumor position is not made firmly for these patients. It can also be assumed that the offset between the two average alignments only coincidentally left the MidV registration with a closer shift to the exRef. However, it is likely that the visual nature of the AIP had something to do with the shift differences.

The results did not show clinically relevant differences in the presence of tumor position change with respect to amplitude. One patient (patient 11) showed a clear, relatively large, and consistent difference between. This calls to question the assumption of using the MidV reference image. However, it's clinical impact can only be investigated through reproducibility in this paper.

$$\Delta exRef/MidV = \sqrt{(LAT_{exRef} - LAT_{MidV})^2 + (AP_{exRef} - AP_{MidV})^2 + (SI_{ExRef} - SI_{MidV})^2} \quad (4)$$

$$\Delta exRef/AIP = \sqrt{(LAT_{exRef} - LAT_{AIP})^2 + (AP_{exRef} - AP_{AIP})^2 + (SI_{ExRef} - SI_{AIP})^2} \quad (5)$$

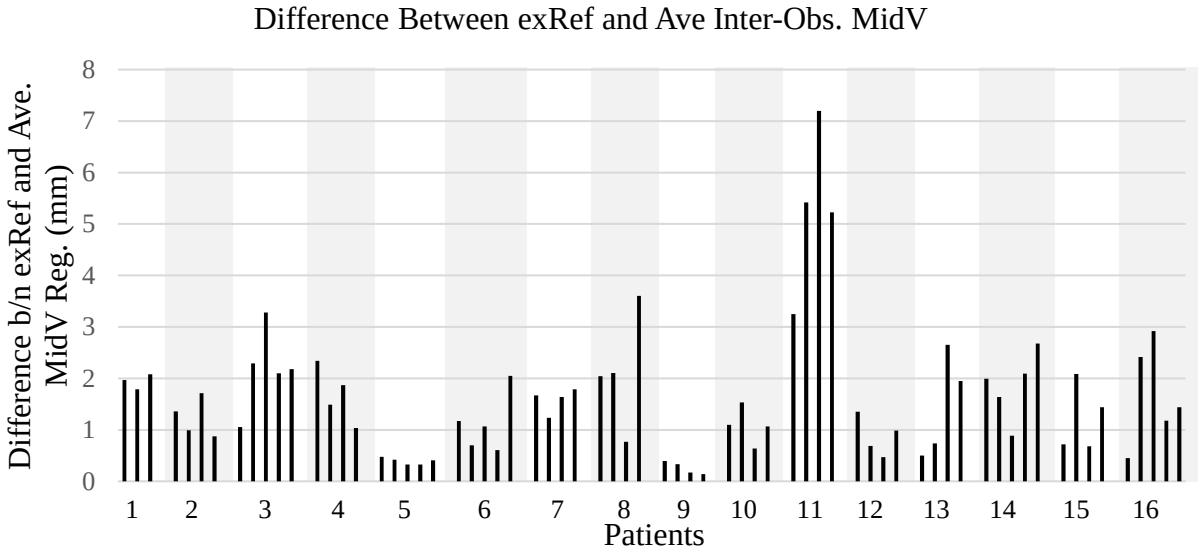

Figure A 19. Shows difference between the shift of the averaged exRef alignments and the average of all MidV inter-observer registrations shifts (Eq. 4). Difference between the two alignments indicates a change in respiratory cycle and/or the misrepresentation of the "most-likely" tumor position at simulation by using the MidV reference image. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background. Patient 11 shows the largest differences between the exRef registration shifts and the MidV registration shifts.

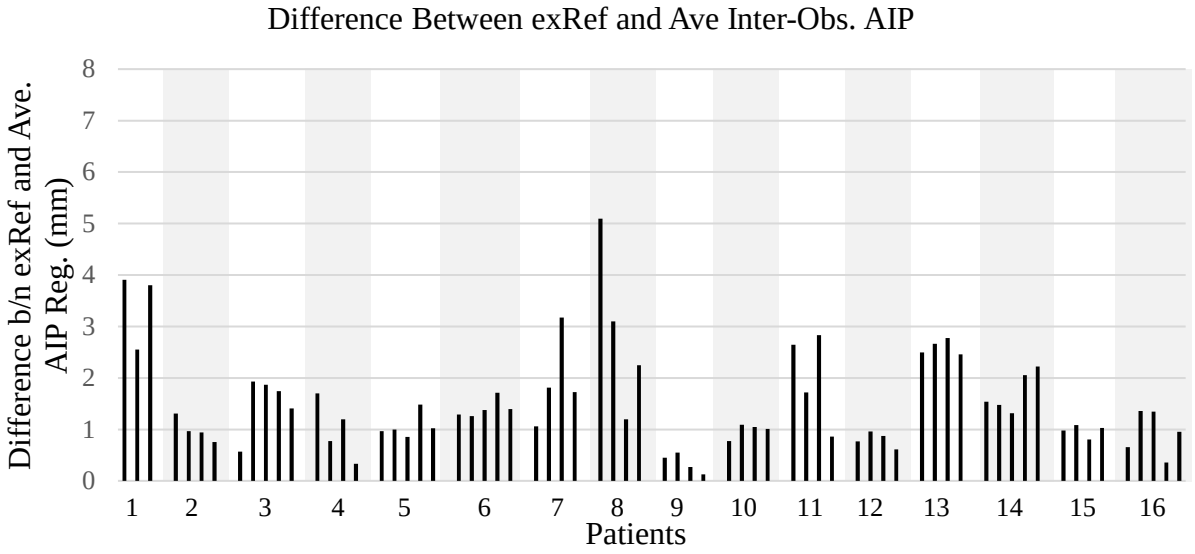

Figure A 20. Shows difference between the shift of the averaged exRef alignments and the average of all inter-observer AIP registrations shifts (Eq. 5). Difference between the two alignments indicates a change in respiratory cycle. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background. Patient 8 shows the largest differences between the exRef registration shift and the AIP registration shift.

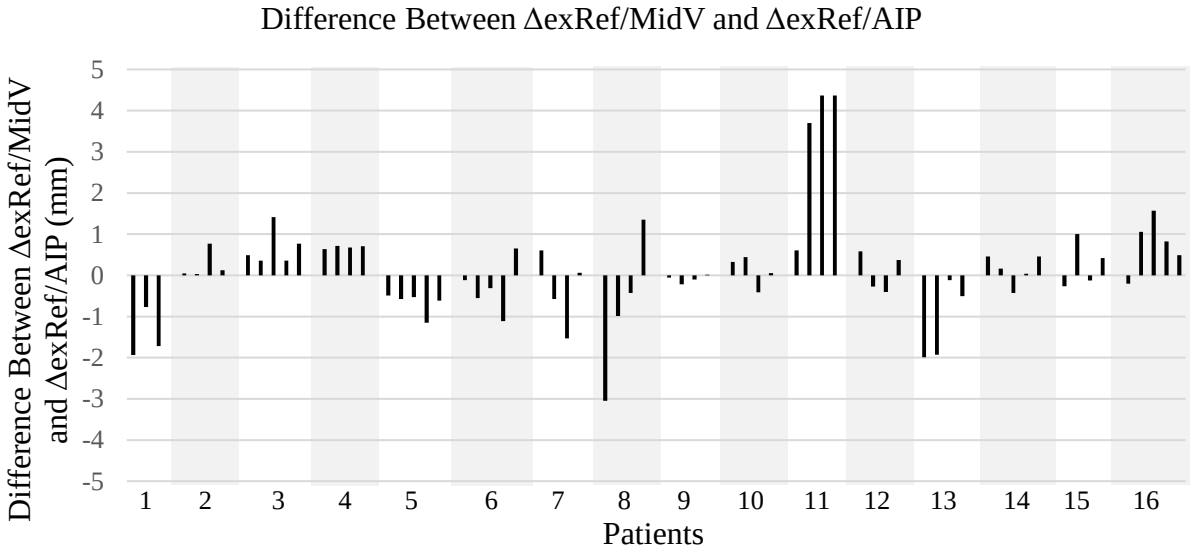

Figure A 21. Shows the difference in results of Fig. A 19 and A 20 (Eq. 4 – Eq. 5). Similar differences (~0 mm) indicate that the breathing cycle did not change or effected both reference images types similarly. Larger differences indicate that one reference image followed the exRef registration more closely than the other. Consistency for all a patient's fractions may indicate a systematic shift. All 16 patients (fractions ranging from 3-5) are shown; patient 1 is the left most set of fractions (white background, three fractions). Every other patient (even numbered patients) are distinguished by a grey background. Odd patients have a white background. Patient 8 shows the largest differences between the exRef registration shift and the AIP registration shift.

Figure A 22 correlates the maximum difference between MidV inter-observer registrations and the difference between the average exRef and average inter-observer MidV registration. The difference between the exRef and average MidV registration quantifies breathing changes between simulation and treatment or, when compared with the AIP results, can also indicate if the MidV was a good representation of the average tumor location during Sim. The maximum inter-observer difference quantifies the maximum range between likely image registrations for a group of observers. Therefore, this plot can indicate how breathing variability or the MidV reference image's incorrect assumptions affects inter-observer registration variability.

Figure A 23 correlates maximum difference between AIP inter-observer registrations and the difference between the average exRef and average inter-observer AIP registration. The difference between the exRef and average MidV registration quantifies breathing changes between simulation and treatment or can potentially indicate a severely blurred AIP image that may not qualify as an appropriate reference image for alignment. The maximum inter-observer difference quantifies the maximum range between likely image registrations for a group of observers. Therefore, this plot can indicate how breathing variability or how the severity of blurriness affects inter-observer registration variability.

Both Figures A 22 and A 23 show weak linear correlations and their results are comparable. Therefore, neither reference image has an advantage when breathing cycle changes are present.

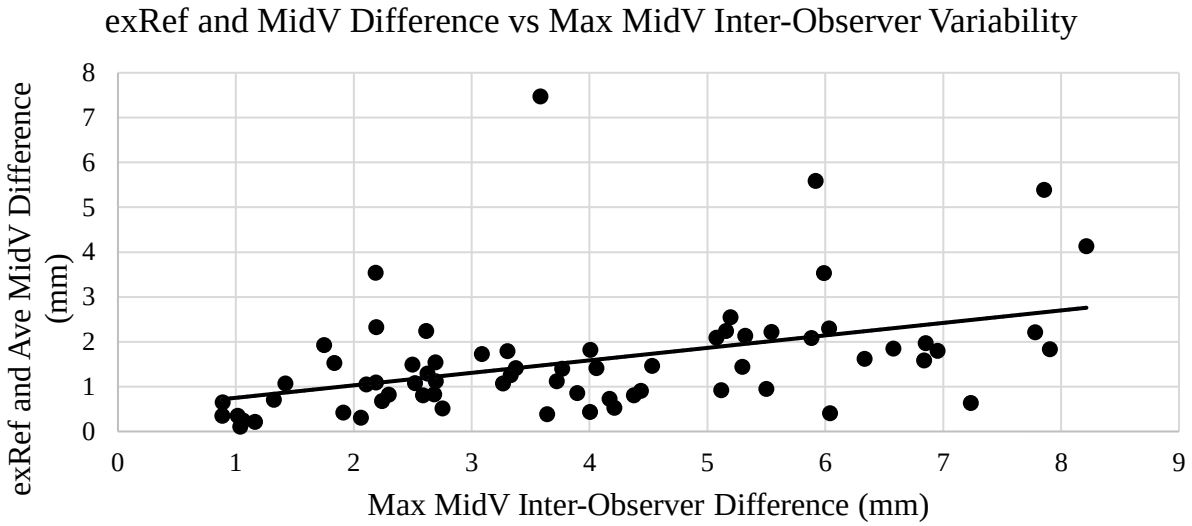

Figure A 182. Scatter plot of the highest difference between inter-observer registrations for the MidV data set (Eq. 3) and the difference between the average exRef and average MidV/CBCT registration (Eq. 4) per fraction. Can indicate how breathing variability or the MidV reference image's lack of correlation with being the most likely tumor position during simulation affects inter-observer registration variability. Shows slight positive trend. R value = 0.428.

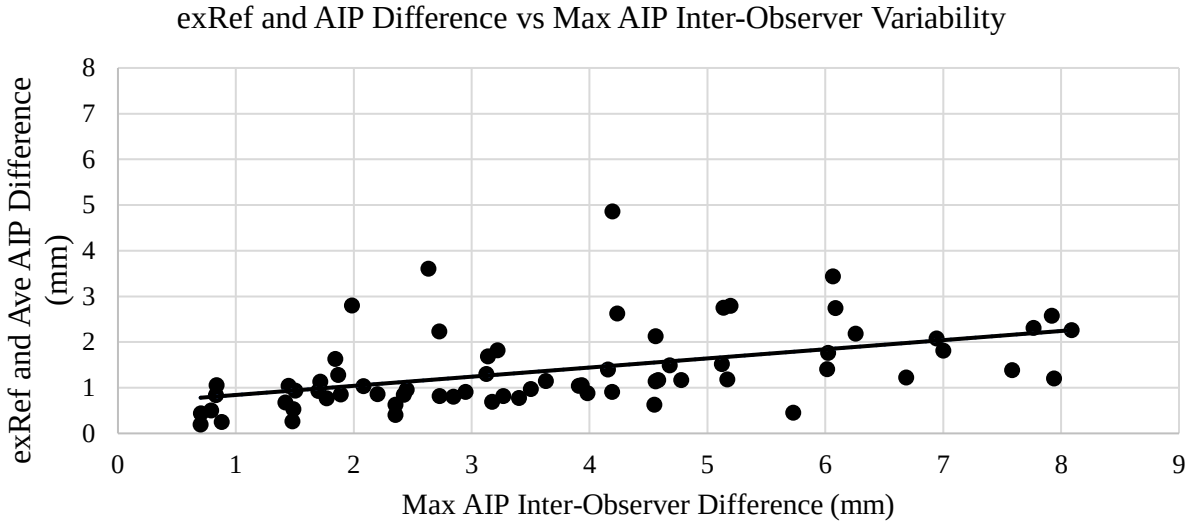

Figure A 193. Scatter plot of the highest difference between inter-observer registrations for the AIP data set (Eq. 3) and the difference between the average exRef and average AIP/CBCT registration (Eq. 5) per fraction. Can indicate how breathing variability affects inter-observer registration variability. Shows slight positive trend. R value = 0.465.

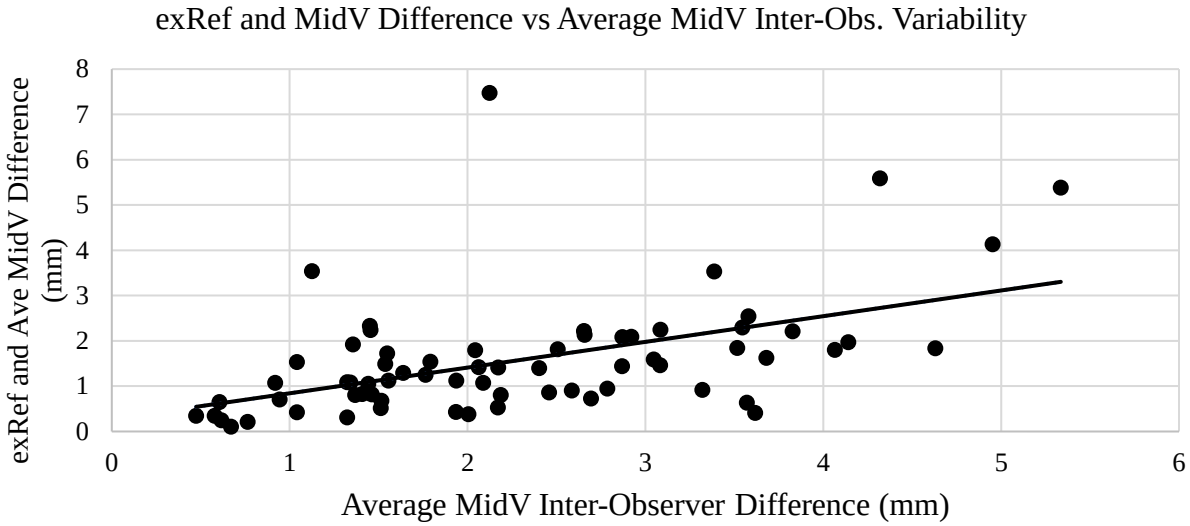

Figure A 24. Scatter plot of the average difference between inter-observer registrations for the MidV data set (Eq. 3) and the difference between the average exRef and average MidV/CBCT registration (Eq. 4) per fraction. Can indicate how breathing variability or the MidV reference image's lack of correlation with being the most likely tumor position during simulation affects inter-observer registration variability. Shows slight positive trend. R value = 0.502.

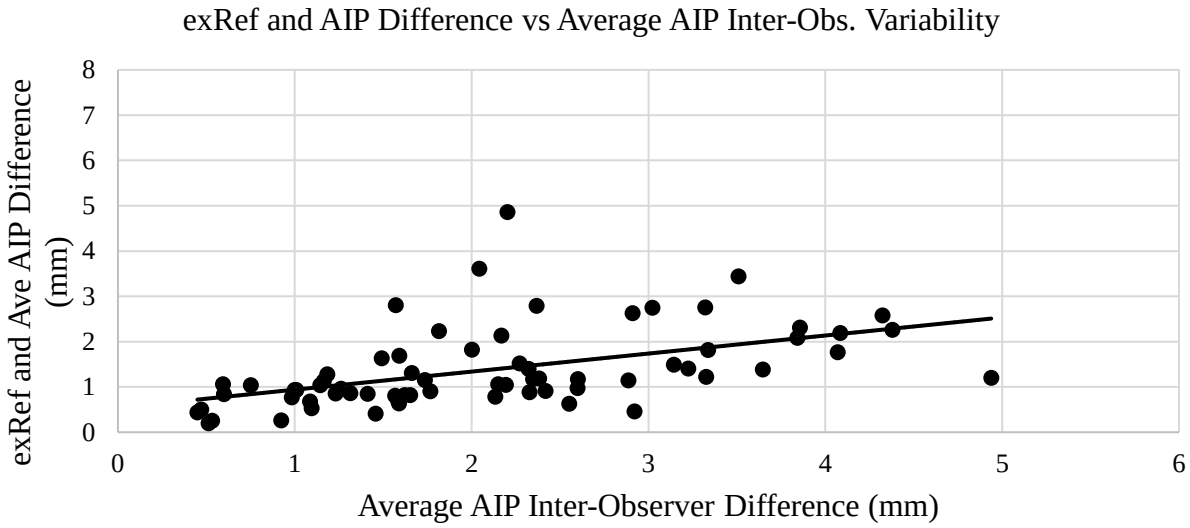

Figure A 205. Scatter plot of the average difference between inter-observer registrations for the AIP data set (Eq. 3) and the difference between the average exRef and average AIP/CBCT registration (Eq. 5) per fraction. Can indicate how breathing variability affects inter-observer registration variability. Shows slight positive trend. R value = 0.492.

References

1. R.L. Siegel, K.D. Miller, and A. Jemal, Cancer statistics, 2019, CA: A Cancer Journal for Clinicians 69(1), 7–34 (2019).
2. J.M. Balter, R.K. Ten Haken, T.S. Lawrence, K.L. Lam, and J.M. Robertson, Uncertainties in CT-based radiation therapy treatment planning associated with patient breathing, International Journal of Radiation Oncology*Biology*Physics **36**(1), 167–174 (1996).
3. The management of respiratory motion in radiation oncology report of AAPM Task Group 76a) - Keall - 2006 - Medical Physics - Wiley Online Library, (n.d.).
4. L. Ekberg, O. Holmberg, L. Wittgren, G. Bjelkengren, and T. Landberg, What margins should be added to the clinical target volume in radiotherapy treatment planning for lung cancer?, Radiotherapy and Oncology **48**(1), 71–77 (1998).
5. D.J. Mohatt, J.M. Keim, M.C. Greene, A. Patel-Yadav, J.A. Gomez, and H.K. Malhotra, An investigation into the range dependence of target delineation strategies for stereotactic lung radiotherapy, Radiat Oncol 12(1), 166 (2017).
6. J.D. Bradley, A.N. Nofal, I.M. El Naqa, *et al.*, Comparison of helical, maximum intensity projection (MIP), and averaged intensity (AI) 4D CT imaging for stereotactic body radiation therapy (SBRT) planning in lung cancer, Radiotherapy and Oncology **81**(3), 264–268 (2006).
7. J.M. Balter, K.L. Lam, C.J. McGinn, T.S. Lawrence, and R.K. Ten Haken, Improvement of CT-based treatment-planning models of abdominal targets using static exhale imaging, International Journal of Radiation Oncology*Biology*Physics **41**(4), 939–943 (1998).
8. L. Bedos, O. Riou, N. Aillères, *et al.*, Evaluation of reproducibility of tumor repositioning during multiple breathing cycles for liver stereotactic body radiotherapy treatment, Rep Pract Oncol Radiother **22**(2), 132–140 (2017).
9. R.W.M. Underberg, F.J. Lagerwaard, B.J. Slotman, J.P. Cuijpers, and S. Senan, Use of maximum intensity projections (MIP) for target volume generation in 4DCT scans for lung cancer, International Journal of Radiation Oncology*Biology*Physics **63**(1), 253–260 (2005).
10. R.W.M. Underberg, F.J. Lagerwaard, B.J. Slotman, J.P. Cuijpers, and S. Senan, Benefit of respiration-gated stereotactic radiotherapy for stage I lung cancer: An analysis of

4DCT datasets, International Journal of Radiation Oncology*Biology*Physics **62**(2), 554–560 (2005).

11. J.D. Bradley, A.N. Nofal, I.M. El Naqa, *et al.*, Comparison of helical, maximum intensity projection (MIP), and averaged intensity (AI) 4D CT imaging for stereotactic body radiation therapy (SBRT) planning in lung cancer, Radiotherapy and Oncology **81**(3), 264–268 (2006).
12. R.W.M. Underberg, F.J. Lagerwaard, J.P. Cuijpers, B.J. Slotman, J.R. van Sörnsen de Koste, and S. Senan, Four-dimensional CT scans for treatment planning in stereotactic radiotherapy for stage I lung cancer, International Journal of Radiation Oncology*Biology*Physics **60**(4), 1283–1290 (2004).
13. S.S. Vedam, P.J. Keall, V.R. Kini, H. Mostafavi, H.P. Shukla, and R. Mohan, Acquiring a four-dimensional computed tomography dataset using an external respiratory signal, Phys. Med. Biol. **48**(1), 45–62 (2002).
14. E.C. Ford, G.S. Mageras, E. Yorke, and C.C. Ling, Respiration-correlated spiral CT: A method of measuring respiratory-induced anatomic motion for radiation treatment planning, Medical Physics **30**(1), 88–97 (2003).
15. J.-P. Bissonnette, T.G. Purdie, J.A. Higgins, W. Li, and A. Bezjak, Cone-beam computed tomographic image guidance for lung cancer radiation therapy, Int. J. Radiat. Oncol. Biol. Phys. 73(3), 927–934 (2009).
16. M. Guckenberger, J. Meyer, J. Wilbert, et al., Cone-beam CT based image-guidance for extracranial stereotactic radiotherapy of intrapulmonary tumors, Acta Oncol 45(7), 897–906 (2006).
17. W. Ottosson, M. Baker, M. Hedman, C.F. Behrens, and D. Sjöström, Evaluation of setup accuracy for NSCLC patients; studying the impact of different types of cone-beam CT matches based on whole thorax, columna vertebralis, and GTV, Acta Oncol 49(7), 1184–1191 (2010).
18. B. Jiang, J. Dai, Y. Zhang, et al., Comparison of setup error using different reference images: a phantom and lung cancer patients study, Med DoSim 37(1), 47–52 (2012).
19. K. Shirai, K. Nishiyama, T. Katsuda, et al., Phantom and clinical study of differences in cone beam computed tomographic registration when aligned to maximum and average intensity projection, Int. J. Radiat. Oncol. Biol. Phys. 88(1), 189–194 (2014).

20. T. Kamomae, H. Monzen, S. Nakayama, et al., Accuracy of image guidance using free-breathing cone-beam computed tomography for stereotactic lung radiotherapy, PLoS ONE 10(5), e0126152 (2015).
21. M. Oechsner, B. Chizzali, M. Devecka, S.E. Combs, J.J. Wilkens, and M.N. Duma, Registration uncertainties between 3D cone beam computed tomography and different reference CT datasets in lung stereotactic body radiation therapy, Radiat Oncol 11(1), 142 (2016).
22. M. Oechsner, B. Chizzali, M. Devecka, et al., Interobserver variability of patient positioning using four different CT datasets for image registration in lung stereotactic body radiotherapy, Strahlenther Onkol 193(10), 831–839 (2017).
23. G.D. Hugo, J. Liang, J. Campbell, and D. Yan, On-line target position localization in the presence of respiration: a comparison of two methods, Int. J. Radiat. Oncol. Biol. Phys. 69(5), 1634–1641 (2007).
24. K. Shirai, K. Nishiyama, T. Katsuda, et al., Maximum Intensity Projection (MIP) and Average Intensity Projection (AIP) in Image Guided Stereotactic Body Radiation Therapy (SBRT) for Lung Cancer, International Journal of Radiation Oncology*Biology*Physics 84(3, Supplement), S718 (2012).
25. M. Oechsner, B. Chizzali, J.J. Wilkens, S.E. Combs, and M.N. Duma, EP-1837: Impact on patient positioning using four CT datasets for image registration with CBCTs in lung SBRT, Radiotherapy and Oncology 119, S862–S863 (2016).
26. C. Garibaldi, G. Piperno, A. Ferrari, et al., Translational and rotational localization errors in cone-beam CT based image-guided lung stereotactic radiotherapy, Phys Med 32(7), 859–865 (2016).
27. N. Wink, C. Panknin, and T.D. Solberg, Phase versus amplitude sorting of 4D-CT data, J Appl Clin Med Phys 7(1), 77–85 (2006).
28. Watkins, W. T., Li, R. , Lewis, J. , Park, J. C., Sandhu, A. , Jiang, S. B. and Song, W. Y. (2010), Patient-specific motion artifacts in 4DCT. Med. Phys., 37: 2855-2861. doi:10.1118/1.3432615
29. Wolthaus JW, Schneider C, Sonke JJ, van Herk M, Belderbos JS, Rossi MM, Lebesque JV, Damen EM. Mid-ventilation CT scan construction from four-dimensional respiration-

correlated CT scans for radiotherapy planning of lung cancer patients. Int J Radiat Oncol Biol Phys. 2006 Aug 1;65(5):1560-71.

30. Seppenwoolde Y1, Shirato H, Kitamura K, Shimizu S, van Herk M, Lebesque JV, Miyasaka K. Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy. Int J Radiat Oncol Biol Phys. 2002 Jul 15;53(4):822-34.
31. Li, G., Citrin, D., Camphausen, K., Mueller, B., Burman, C., Mychalczak, B., Miller RW, Song, Y. (2008). Advances in 4D Medical Imaging and 4D Radiation Therapy. Technology in Cancer Research & Treatment, 67–81.