

5-2014

EVALUATION OF ARTIFACTS IN EXPERIMENTAL CINE 4D CT ACQUISITION METHODS

Sarah Joy

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations



Part of the [Medical Biophysics Commons](#), and the [Other Physics Commons](#)

Recommended Citation

Joy, Sarah, "EVALUATION OF ARTIFACTS IN EXPERIMENTAL CINE 4D CT ACQUISITION METHODS" (2014). *The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access)*. 431.
https://digitalcommons.library.tmc.edu/utgsbs_dissertations/431

This Dissertation (PhD) is brought to you for free and open access by the The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digitalcommons@library.tmc.edu.

EVALUATION OF ARTIFACTS IN EXPERIMENTAL CINE 4D CT
ACQUISITION METHODS

by

Sarah Joy Castillo, M.S.

APPROVED:

Thomas Guerrero, M.D., Ph.D.
Supervisory Professor

Richard Wendt, Ph.D.

Tinsu Pan, Ph.D.

Geoffrey Ibbott, Ph.D.

Brian Hobbs, Ph.D.

APPROVED:

Dean, The University of Texas
Health Science Center at Houston
Graduate School of Biomedical Sciences

EVALUATION OF ARTIFACTS IN EXPERIMENTAL 4D CT
ACQUISITION METHODS

A

DISSERTATION

Presented to the Faculty of
The University of Texas
Health Science Center at Houston
and
The University of Texas
MD Anderson Cancer Center
Graduate School of Biomedical Sciences
in Partial Fulfillment
of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Sarah Joy Castillo, M.S.

Houston, Texas

May 2014

ACKNOWLEDGEMENTS

I would like to thank my advisory committee for their guidance through this degree, particularly my advisor Thomas Guerrero, M.D., Ph.D., for the opportunity and support for this research through the NIH Director's New Innovator Award (DP2OD007044). The members of the consensus group have also been vital to the success of the first research aim, thank you Dr. Guerrero, Nicholas Murray, Adenike Olanrewaju, Dr. Sastry Vedam, and Dr. Julie Pollard. Dr. Brian Hobbs, thank you very much for the statistical support.

I would like to thank David Leos for being so accommodating with patient recruitment for the protocol, and for helping make the process as efficient as possible. The physics assistants and therapists have been of great aid in assisting with the protocol process, and I thank them greatly for their patience and assistance. Jared Ohrt and Mike Kantor have also been extremely helpful throughout my graduate career, and I am grateful for their assistance. My greatest thanks goes to the patients who participated in this study; their noble desire to help others in their situation allows advancement in the field for future generations.

Dr. Richard Wendt and Dr. Rebecca Howell have been incredibly supportive throughout my graduate career and have graced me with advice and aid in times of need, and I am truly grateful for their relationship. I am also very grateful to the wonderful GSBS faculty for their support, especially Dr. William Mattox and Brenda Gaughan, and Dr. Victoria Knutson. I would also like to acknowledge Dr. Edward Jackson for his amazing support and guidance, not to just to me, but to all the physics students.

Thank you to all the administrative supporters! Jessica Hunter, Karen Lugo, Betsy Kindred, Lisa Watson, Norma Hall, and Sarah Welch.

I am very thankful for help provided by my brother-in-law Dr. Eddie Castillo, in enabling an efficient implementation of our experimental sorting method, and for moral support throughout this degree.

Finally, I am indebted to my husband Dr. Richard Castillo for providing vital software, teaching me programming skills, assistance with editing, advice, help with helical 4D CT, and never ending moral and emotional support. I could not have done this without you.

EVALUATION OF ARTIFACTS IN EXPERIMENTAL 4D CT ACQUISITION

METHODS

by

Sarah Joy Castillo, M.S.

Supervisory Professor: Thomas Guerrero, M.D., Ph.D.

Four-dimensional computed tomography (4D CT) has increased the accuracy of radiation treatment planning for patients in whom the extent of target motion is large. 4D CT has become a standard of care for radiation treatment simulation, allowing decreased motion artifacts and increased spatiotemporal localization of anatomical structures that move. However, motion artifacts may still remain. These artifacts, or artificial anatomic spatial distributions, add a systematic uncertainty to the treatment process and limit the accuracy of lung function images derived from CT. We proposed to reduce the motion artifacts in cine 4D CT by using three novel investigational 4D CT acquisition methods: (1) oversampling the data acquired, (2) gating the x-ray beam with breathing irregularities, and (3) rescanning areas of the clinical standard 4D CT associated with high breathing irregularities. These experimental acquisitions were tested through a protocol approved by the institutional review board with 18 patients with a primary thoracic malignancy receiving a standard 4D CT scan for radiation treatment simulation. The artifact presence in all 4D CT scans was assessed by an automated artifact quantification metric. This artifact metric was validated by a rigorous receiver operating characteristic (ROC) analysis using a high-quality dataset derived from a group of expert observers who

reached a consensus decision on the artifact frequency and magnitude for each of 10 clinical 4D CT scans from patients with primary thoracic cancer. The clinical and experimental 4D CT acquisitions from the 18 patients on the protocol were post-processed by the clinical standard of phase sorting and by an experimental phase sorting that incorporated the validated artifact metric. The 4D CT acquisition and processing method judged to be the most improved was the oversampling acquisition with the experimental sorting. The reproducibility of this improved method was tested on a second distinct cohort of 10 patients with a primary thoracic malignancy. Those patients received a clinical phase-sorted 4D CT immediately followed by three independent oversampling acquisitions, processed by the experimental sorting method and evaluated using the artifact metric. The experimental-sorted oversampling acquisition produced a statistically significant artifact reduction (27% and 28% per cohort) from the phase-sorted clinical standard acquisition.

TABLE OF CONTENTS

CHAPTER 1: BACKGROUND & INTRODUCTION

MOTIVATION	1
4D CT PROCESS.....	2
I. Respiratory Characterization	3
II. 4D CT Modes	4
i. <i>Helical 4D CT</i>	4
ii. <i>Cine 4D CT</i>	5
4D CT ARTIFACTS.....	7
I. Artifact Definition	8
II. Artifact Classification	8
i. <i>Helical vs. Cine Artifact</i>	8
ii. <i>Yamamoto Classification</i>	10
iii. <i>Cardiac Artifacts</i>	11
4D CT ARTIFACT QUANTIFICATION	11
I. Manual Assessment	12
II. Quantitative Assessment.....	12
4D CT ARTIFACT REDUCTION.....	14
I. Alternative Binning Methods	15
i. <i>Amplitude Binning</i>	15
ii. <i>Correlation-Based Processing</i>	18
II. Registration Methods.....	19
III. Alternative Acquisition Methods	20
HYPOTHESIS & SPECIFIC AIMS	22

CHAPTER 2: VALIDATION OF THE ARTIFACT QUANTIFICATION METRIC

INTRODUCTION.....	24
--------------------------	-----------

THE ARTIFACT QUANTIFICATION METRIC.....	25
I. The Normalized Cross Correlation Coefficient.....	25
II. The Pearson Correlation Coefficient.....	25
III. The Final Artifact Metric	26
THE SAMPLE EVALUATED	27
I. The Sample Data	27
II. The Sample Size Calculation	28
OBSERVER ARTIFACT ASSESSMENT	29
I. Consensus Group.....	29
II. Consensus Group Instructions	29
III. Artifact Evaluation Method	30
IV. Analysis of Consensus-Chosen Artifacts	32
ARTIFACT THRESHOLD DERIVATION	33
I. ROC Curve Generation	33
II. ROC Curve Parameters.....	34
III. Artifact Threshold.....	34
ARTIFACT THRESHOLD EVALUATION	35
METRIC VS CONSENSUS-CHOSEN ARTIFACT MAGNITUDES	35
BREATHING IRREGULARITY.....	35
I. Breathing Irregularity Identification	35
II. Breathing Irregularity vs. Consensus-Chosen Artifacts.....	36
RESULTS	37
I. The Artifact Metric	37

II. The Consensus-Chosen Artifacts	39
III. ROC Analysis	41
IV. Metric vs Consensus-Chosen Artifact Magnitudes.....	42
V. Breathing Irregularities vs. Consensus-Chosen Artifacts	43
DISCUSSION	45
CONCLUSION	49

CHAPTER 3: EXPERIMENTAL 4D CT ACQUISITIONS

INTRODUCTION.....	50
I. The Patient Sample	51
II. 4D CT Parameters	52
III. Experimental Methods	53
i. <i>Gating Method</i>	<i>53</i>
ii. <i>Rescan Method.....</i>	<i>53</i>
iii. <i>Oversampling Method.....</i>	<i>53</i>
iv. <i>Breathing Irregularity Detection</i>	<i>54</i>
v. <i>Acquisition Order.....</i>	<i>54</i>
IV. Post-Processing Methods.....	55
i. <i>Phase Sorting.....</i>	<i>55</i>
ii. <i>Experimental Sorting</i>	<i>56</i>
V. Artifact Evaluation	57
i. <i>Artifact Analysis.....</i>	<i>57</i>
ii. <i>Statistical Analysis.....</i>	<i>57</i>
VI. Effective Dose Estimates.....	58
VII. Artifact Presence vs Breathing Irregularities	58
RESULTS	58
I. Patient Summary Statistics	58
II. Experimental Acquisition Rankings.....	59

i. <i>Phase-Sorted Acquisition</i>	59
ii. <i>CM-Sorted Acquisition</i>	60
III. Statistical Analysis	63
i. <i>Phase-Sorted Acquisition</i>	63
ii. <i>CM-Sorted Acquisition</i>	63
iii. <i>CM vs Phase Sorting</i>	63
IV. Effective Dose Estimates	64
V. Breathing Irregularities vs. Artifact Presence	65
DISCUSSION	69
CONCLUSION	73

CHAPTER 4: BEST 4D CT METHOD REPRODUCIBILITY

INTRODUCTION	74
I. The Patient Sample	74
II. Improved Experimental Method	74
III. Analysis	75
i. <i>Artifact Presence</i>	75
ii. <i>Image Quality vs. Patient Parameters</i>	76
IV. Effective Dose Estimates	76
RESULTS	76
I. Patient Respiratory Statistics	76
II. Analysis of Artifact Presence	77
III. Image Quality vs. Patient Parameters	83
IV. Effective Dose Estimates	83
DISCUSSION	84
CONCLUSION	87
REFERNCES	89

List of Illustrations

Figure 1.1 Cine 4D CT Artifact Example	9
Figure 1.2. Helical 4D CT Artifact Example	9
Figure 1.3. Yamamoto Artifact Classification	10
Figure 1.4. Cardiac Artifact Example	11
Figure 2.1. Artifact Evaluation Software	31
Figure. 2.2. Breathing Irregularity Identification.....	36
Figure 2.3. Normalized Correlation Metric vs Couch Positions, Patient C2_1	38
Figure 2.4. Example Coronal Slices of Patient C2_1	39
Figure 2.5. Estimated Probability of Artifacts as a Function of Phase	40
Figure 2.6. Example ROC Curve from C1_5.....	41
Figure 2.7. Percentage Breathing Irregularities at Artifact Locations	43
Figure 2.8. Percentage Artifacts at Irregular Breath Locations	44
Figure 2.9. Artifact Scores and Breathing Irregularities.....	45
Figure 3.1. Phase-Sorted Acquisition Rankings	60
Figure 3.2. CM-Sorted Acquisition Rankings	61
Figure 3.3. CM-Sorted Sample Coronal Views	62
Figure 3.4. Sample Coronal Views for Sorting Comparison	62
Figure 3.5. Estimated Comparable Effective Doses	64
Figure 3.6. Phase-Sorted Artifacts vs Number of Breathing Irregularities.....	65
Figure 3.7. CM-Sorted Artifacts vs Number of Breathing Irregularities.....	66
Figure 3.8. Phase-Sorted Artifacts vs Percentage of Breathing Irregularities	67
Figure 3.9. CM-Sorted Artifacts vs Percentage of Breathing Irregularities	68

Figure 3.10. Gating Scan Duration vs Breathing Irregularities	69
Figure 4.1. Boxplot of CM Values by Acquisition Method.....	78
Figure 4.2. Bland-Altman Plot for Oversampling CM Value Inter-Acquisition Agreement.....	79
Figure 4.3. CM Value Spread per Patient.	80
Figure 4.4. Coronal views of the clinical and repeat OCM scans.....	81

List of Tables

Table 2.1	Cohort 1 ROC Parameters.....	42
Table 2.2	Cohort 2 Sensitivity and Specificity Values	42
Table 4.1.	Cohort 2 Patient Respiratory Statistics	77
Table 4.2.	Spearman correlation coefficients between CM values and breathing parameters	82
Table 4.3.	Spearman correlation between percentage improvement in CM values with OCM and percentage difference in breathing parameters	82

List of Equations

Equation 1.1	Helical Pitch	5
Equation 1.2	Image Segments per Cine Couch Position	6
Equation 2.1	Normalized Cross Correlation Coefficient.....	25
Equation 2.2	2D Pearson Correlation Coefficient.....	26
Equation 2.3	Correlation Metric	27
Equation 2.4	Exact Test for Single Proportion.....	28
Equation 2.5	Youden Index	34

List of Abbreviations

4D CT	Four-dimensional computed tomography
3D CT	Three-dimensional computed tomography
RPM	Real-time processing monitor
GTV	Gross tumor volume
NCC	Normalized correlation coefficient
CM	Correlation metric
ROC	Receiver operating characteristic
AUC	Area under the curve
SD	Standard deviation
GM	Gating acquisition method
RM	Rescan acquisition method
OM	Oversampling acquisition method
CTDI	Computed tomography dose index
ANOVA	Analysis of variance
OCM	Oversampling acquisition with correlation metric sorting

CHAPTER 1

BACKGROUND & INTRODUCTION

MOTIVATION

Four-dimensional computed tomography (4D CT) is routinely employed as an integral part of radiation therapy simulation when there is a need to account for respiratory motion.⁽¹⁻⁷⁾ 4D CT relates the image acquisition with the patient's breathing resulting in a series of 3D image volumes that represent the breathing cycle from peak inhalation through exhalation to peak inhalation.^(8, 9) This method enables more accurate treatment delivery by limiting the uncertainty in the target cancer location as the target deforms with respiratory motion.^(5, 10, 11) However, the ability of the 4D CT to limit uncertainties associated with anatomic position depends on image quality.

Artifacts, or artificial anatomic spatial distributions, cause uncertainty in the true anatomic spatial distribution that lead to errors in anatomic delineation, treatment targeting, and lung function images derived from CT. 4D CT may contain both standard 3D CT artifacts and those specific to the 4D acquisition and mode. Artifacts specific to 4D CT often arise from breathing irregularities that are currently not accounted for in the clinical setting. These irregularities may introduce appreciable artifacts into the images, and the only recourse to using the low-quality dataset for treatment planning is to re-acquire a clinical 4D CT scan. If the re-acquired 4D CT demonstrates appreciable artifacts, free-breathing images that do not contain these artifacts are used for treatment planning. These images do not account for respiratory motion, and represent multiple breathing states within one 3D CT.

To the best of our knowledge, methods to prospectively reduce 4D CT image artifacts by altering the acquisition have not been attempted in a clinical setting. In the present study, we implemented three experimental 4D CT acquisition methods that target breathing irregularities (acquiring more images, gating the x-ray beam with breathing irregularities, and re-acquiring images associated with breathing irregularities) in patients with thoracic cancer to determine the potential for improving 4D CT image quality in a clinical setting in a relatively simple manner. Unlike in most previous studies, the methods we used focused on acquisition modification rather than retrospective analysis.

4D CT PROCESS

4D CT is a common motion-management method used in radiation treatment simulation. The American Association of Physicists in Medicine (AAPM) recommends the use of a motion management simulation method in the presence of anatomic motion that exceeds 5 mm in any direction,⁽⁶⁾ which occurs in about 40% of lung targets.⁽¹²⁾ Because thoracic and some abdominal structures meet this motion threshold, 4D CT has become the standard protocol for patients scheduled to undergo radiation treatment for thoracic cancer at The University of Texas MD Anderson Cancer Center.

The 4D CT images are used as the primary dataset for the radiation treatment plan; target and normal tissue contours are delineated and defined on the primary dataset. Radiation fields and beam arrangements are designed based on the contours to achieve target coverage and normal tissue dose constraints. Anatomic uncertainty arising from the 4D CT primary dataset will propagate into the treatment plan and delivery as a systematic uncertainty and will increase the difference between the planned and the delivered dose.

I. Respiratory Characterization

4D CT typically relies on an external motion surrogate to capture breathing information that is independent of the image acquisition. In the most common setup, an abdominal belt or a real-time processing monitor (RPM) box is placed on top of the patient's abdomen between the xiphoid process and umbilicus. The surrogate records a 1-dimensional signal of either relative mechanical pressure (belt) from a pressure sensor or relative abdominal height from infrared reflective dots on the box (RPM); this 1D signal represents the patient's respiratory trace. These two external surrogates have found to be equivalent systems in terms of image quality and effect on treatment planning,⁽¹³⁻¹⁵⁾ but they have also been found to exhibit good but not exact correlation with internal anatomic motion.⁽¹⁶⁻¹⁹⁾

A 4D CT scan comprises a set of 3D CT scans; each 3D CT image set represents a phase of the breathing cycle as defined from peak inhalation through exhalation to peak inhalation on the external surrogate respiratory trace. Thus each voxel must be imaged for at least one breathing cycle or phase representations will be inaccurate. Peak inhalation positions are first computationally defined by smoothing the waveform and finding the local maxima; then they are reviewed by the user and re-defined if a false peak was computationally selected.

Breathing phases may be defined by even divisions in the breathing period (time between peaks of a cycle), by five even divisions in time between the initial peak inhalation and the peak exhalation and five even divisions in time between peak exhalation and the final inhalation, or by five even divisions in abdominal

displacement between the initial peak inhalation and the peak exhalation and five even divisions in abdominal displacement between peak exhalation and the final inhalation. The number of breathing phases per cycle is user-defined, but typically set at 10. Generating fewer breathing phases may ensure non-overlapping data and quicker delineation times while generating more breathing phases leads to potential overlap in data (using the same images in multiple phases) and longer delineation times. The choice of 10 phases also derives from sampling requirements, as a typical breathing cycle lasts five seconds and current 4D-capable CT scanners have roughly a 0.5 second tube rotation time.

Secondary image sets are often derived from the 4D CT breathing phases; most commonly derived are the maximum intensity projection (MIP), a 3D CT consisting of the maximum pixel intensities present in the phases, and the average (AVG), composed of the mean pixel intensities present in the phases.

II. 4D CT Modes

Two modes of 4D CT are currently clinically available: helical and cine. The breathing aspects of 4D CT remain the same for each implementation, but the image acquisition and post-processing differ significantly.

i) Helical 4D CT

Helical CT is currently the most commonly used mode of 3D CT in diagnostic imaging because of its high efficiency and the ability to reconstruct images at any point along the patient's cranial-caudal (\hat{z}) direction. Slip rings enable the efficiency by allowing a continuous x-ray tube rotation for the scan

duration. The reconstruction ability derives from how the data are acquired; the x-ray tube is rotating while the table below the patient moves, creating a helical photon projection. Data are then interpolated so that a continuous data set exists, allowing image reconstruction at any \hat{z} plane. The pitch is defined as the table motion per 360 degree tube rotation, F_{table} divided by the collimated beam width, nT (Eqn. 1.1). Typical diagnostic pitch values range from 0.75 to 1.5, with 1.0 allowing data acquisition for full anatomy without overlap.⁽²⁰⁾

$$pitch = \frac{F_{table}}{nT} \quad (1.1)$$

In 4D CT, the acquisition process remains the same, but the pitch is lowered, typically less than 0.1, so that each voxel is imaged for at least one breathing cycle.⁽²¹⁻²³⁾ Images are reconstructed for all breathing phases and binned. The interpolation and continuous acquisition of data in helical 4D CT results in artifacts that seem to vary more smoothly (Figure 1.2) in a particular location compared with cine artifacts. The appearance of lines will often occur if a breathing irregularity occurred and a voxel was not sampled for a full breathing cycle.⁽²³⁾

ii) Cine 4D CT

Cine 4D CT derives from the axial mode of diagnostic CT, which acquires images in collimated beam width size segments along the cranial-caudal direction while the table is stationary. This “step-and-shoot” method requires a cine duration input rather than a pitch to specify the time duration of image

acquisition at each couch step for data sufficiency. At MD Anderson Cancer Center the cine duration is calculated by assessing the patient's breathing with the external surrogate and adding one second to the average breathing period. The extra second ensures the data acquisition sufficiency condition is met. The data sufficiency condition requires the acquisition of data at each anatomic location to be one breathing cycle plus the time needed to acquire one image, which is equal to a gantry rotation. The gantry rotation used in this study is 500 ms, and an extra 500 ms is added to the cine duration to further ensure the data sufficiency condition is met if the patient breathes more rapidly during image acquisition. The cine time between images represents how often in time an image segment is reconstructed. Cine 4D CT reconstructs images in real-time while helical reconstructs the images after acquisition.

To calculate the number of image segments per couch position along the cranial-caudal direction, the amount of time needed for data acquisition required to form the initial image segment is subtracted from the cine duration, the result is then divided by the cine time and 1 is added (Eqn. 1.2). For example, if the initial image segment time is 500 ms per revolution, with a 250 ms cine time and 6s cine duration, the total image segments per couch position will be 23, with the first segment's acquired data complete 500 ms after initial beam-on and the second segment's acquired data 750 ms after initial beam-on.

Number of Image Segments

$$= \frac{(\text{cine duration} - \text{time to acquire 1st segment})}{\text{cine time}} + 1$$

The initial projection time depends on whether the reconstruction is a full reconstruction that requires 360 degrees of data for image segment reconstruction or a segmental reconstruction of two thirds of 360 degrees (240 degrees) of data required for each image segment reconstruction. The segmental reconstruction increases temporal resolution at the expense of higher noise and image artifacts; therefore the full reconstruction is standard at MD Anderson.

Because the cine image segments are acquired discretely and independently of the external surrogate, the images cannot be reconstructed at defined phases. Instead their occurrence in time is matched to the closest phase definition in a nearest-neighbor technique termed phase sorting. Phase sorting is the clinical standard post-processing method, which uses time both to define phases based on placement within a user-defined breathing period and to select the ‘nearest neighbor.’ Amplitude sorting is another common technique, but is not commercially available; the T50%, or sixth phase definition is pushed to exhalation and remaining phases are linearly defined based on amplitude to inhalation. This process leads to artifacts that exhibit discontinuous banding that typically appear medially across the anatomy (e.g., Figure 1.1). All 4D CT scans used in this research were acquired with a full reconstruction, with a detector configuration of 8 x 2.5 mm slices, with a cine time between images ranging between 350 ms and 450 ms.

4D CT ARTIFACTS

I. Artifact Definition

Ideally the spatial distribution of pixel intensities of a physical object being imaged exactly match the spatial distribution of pixel intensities of the image; an artifact reflects a disruption in the correspondence between these intensities.^(24, 25) An artifact represents the true physical anatomy inaccurately in the image and may manifest in a variety of ways depending on the cause; the artifacts we focus on in this research are those arising from the cine 4D CT acquisition and processing methods.

II. 4D CT Artifact Classifications

i) Helical vs. Cine Artifacts

As discussed in Chapter 1 section II, helical and cine 4D CT artifacts differ based on the acquisition and processing used. Cine 4D CT artifacts exhibit a banding discontinuity that typically appears medially across the entire sagittal or coronal view (Figure 1.1), whereas helical 4D CT artifacts tend to exhibit more smooth and localized (Figure 1.2) artifacts on the sagittal or coronal views. Helical 4D CT artifacts may also exhibit banding discontinuities across an image slice in the presence of large irregularities.

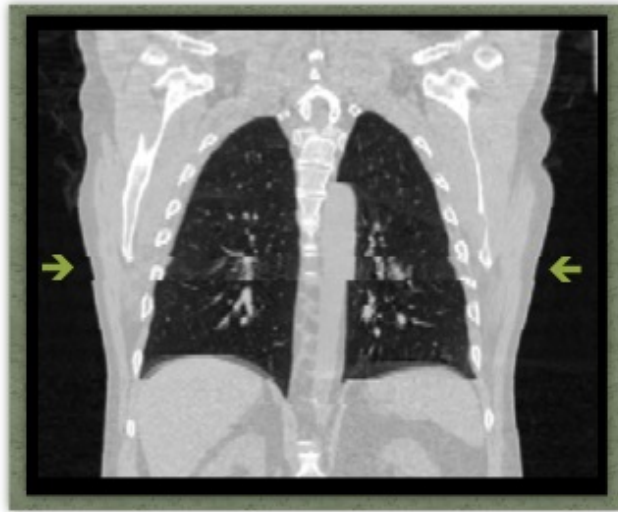


Figure 1.1. Cine 4D CT Artifact Example. Green arrows point to a couch position containing a cine artifact.

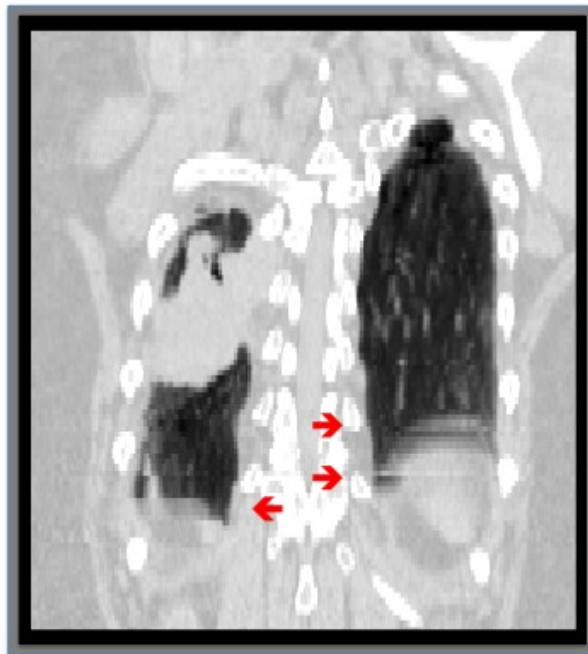


Figure 1.2. Helical 4D CT Artifact Example. Red arrows point to areas containing artifacts.

ii) Yamamoto Classification

A study by Yamamoto et al.⁽²⁶⁾ divided artifacts seen in 4D CT images into four categories: blurring, duplicate, overlapping, and incomplete (Figure 1.3). Blurring, or a partial projection artifact, occurs when organ motion is faster than the speed of the image acquisition; this does not derive from the 4D process but rather from patient-specific motion and scanner speed limitations.⁽²⁷⁾ The other three classifications of artifacts derive from the 4D method and are common, with reportedly roughly equal probability of occurrence and no significant difference in artifact magnitude for each type.

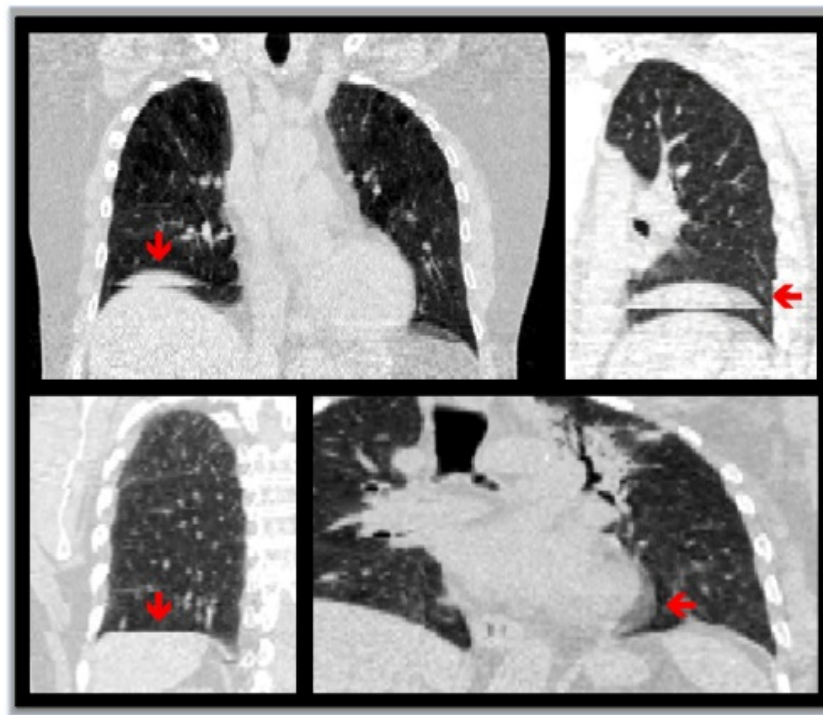


Figure 1.3. Yamamoto Artifact Classification. Overlap (top left), duplicate (top right), incomplete (bottom left), and blurring (bottom right).

iii) Cardiac Artifacts

Cardiac artifacts appear in 4D CT scans as well and derive from the heart beating asynchronously with breathing motion (Figure 1.4). Because cardiac artifacts are not correctable with breathing, they will not be considered in this study.

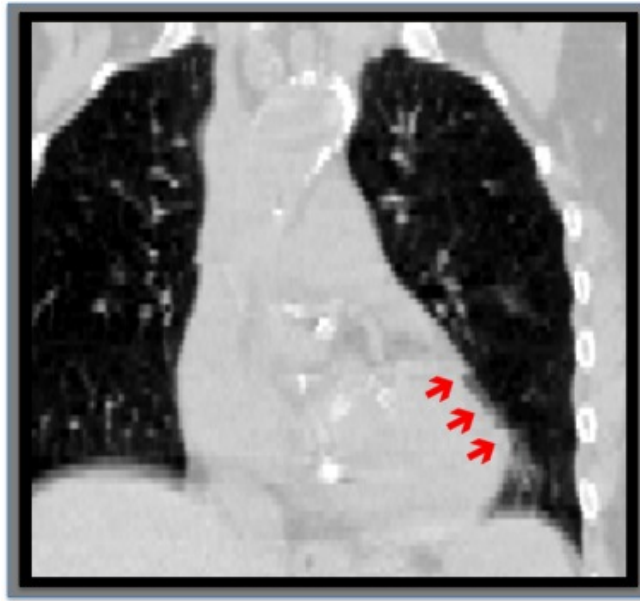


Figure 1.4. Cardiac Artifact Example. Red arrows point to the cardiac artifact.

4D CT ARTIFACT QUANTIFICATION

After a patient with thoracic cancer receives a 4D CT at MD Anderson, a decision is made whether or not to acquire a second 4D CT scan based on the extent of artifact presence in the first 4D CT. The second 4D CT is then evaluated, and if neither scan is acceptable for treatment planning, then the free-breathing helical scan acquired over multiple breathing states is used as an anatomic reference. The evaluation involves manually assessing the overall acceptability of artifacts; this method is subjective,

inconsistent, and requires the physician, physicist, or physics assistant to manually evaluate the scan, which requires time and attendance. Currently some form of manual observation is the standard of artifact evaluation in the clinic and in research studies, however a reliable and efficient evaluation method is warranted.

I. Manual Assessment

Manual assessment by expert observers remains the standard method of artifact evaluation but lacks guidelines, which makes artifact quantification difficult. Some investigators display sample images with a general statement of 4D CT scan quality in a relative fashion,^(1, 28-40) and others define more explicitly a method of manual evaluation by independent experts.^(26, 41-44) Manual artifact assessment is also subjective, with potentially high inter-observer variation, and lengthy analysis times.⁽⁴⁵⁾ Methods of manual assessment vary between institutions, making comparisons of evaluated datasets between groups difficult. A robust, reproducible, automated evaluation method that multiple centers can implement would be ideal, although distinguishing true anatomic changes from false changes resulting from 4D CT presents a challenge. The issues surrounding standard manual assessment also make validation of the accuracy of quantitative methods difficult.

II. Quantitative Assessment

A limited number of studies focus on a metric of artifact evaluation; most studies explore routes of artifact reduction with a simple method of quantifying the reduction. Given the lack of artifact quantification methods, many studies have used

images derived from phantoms and software simulations, in which the exact anatomic volume and shape are known.^(30, 35-37, 46-48)

Persson et al.^(49, 50) evaluated artifacts in terms of gross tumor volume (GTV) deviations from a reference target volume. This method is simple and the GTV is delineated routinely on 4D CT scans used in treatment planning, however it does not reflect spatial deviations or normal anatomic variations and may severely underestimate the frequency of artifacts that are present. Other simple quantitative artifact evaluations have included deviations in target centroid position or contours^(39, 51, 52) (similar to the Persson study), the mean square gray value difference between couch positions,⁽¹⁾ tidal volume variations from reference images,⁽³⁸⁾ and external surrogate parameters.⁽⁵³⁾ These evaluation methods do not account for normal anatomic variation and have not been validated against the standard manual assessment. There are also studies that include only a general statement that the quantitative results match observer results.^(1, 39, 54)

Han et al.⁽⁵⁵⁾ used a normalized cross-correlation coefficient (NCC) to evaluate artifacts in five helical 4D CT scans. A 'bridge stack' was identified as a multi-slice set of images that overlapped locations with two adjacent multi-slice image sets from the same phase, but occurred at a different phase than the two adjacent images. No single threshold was found for NCC values in artifact identification, so each patient had a range of sensitivity and specificity values, making it difficult to determine the efficacy of this evaluation method. The standard against which the NCC values were compared was not derived from manual assessment but instead from the respiratory trace acquired by an external motion

surrogate. The external surrogate does not correlate exactly with internal motion or artifacts and therefore should not be substituted for actual image evaluation. This method also is applicable only to helical 4D CT, and may be degraded if artifacts are present in the reference bridge stack.

Cui et al.⁽⁵⁶⁾ devised a correlation metric (CM) for the assessment of cine artifacts that attempts to account for normal anatomic variations. This metric correlates the image slices across two neighboring couch positions and subtracts the average of correlations between adjacent image slices within the neighboring couch positions. This is a straightforward, efficient metric that could distinguish normal anatomic variation from artifact and is very promising. However, this metric's validation was limited to relative metric values between 10 4D CT scans with two observers.

Because an accurate method of artifact evaluation has not been established yet remains critical for reduction studies, we rigorously validated the accuracy of the Cui metric⁽⁵⁶⁾ which we believe to be the most promising metric.

In this work, a consensus group evaluated artifacts in 10 4D CT scans to form a standard dataset. This dataset was then used to assess the performance of the correlation metric in artifact identification⁽⁵⁶⁾ through receiver operating characteristic (ROC) analysis.

4D CT ARTIFACT REDUCTION

Several efforts to reduce 4D CT artifacts have been undertaken but currently none has been proven to significantly reduce artifacts. These efforts are generally in three main

categories: alternative binning techniques, registration techniques, and alternative acquisition techniques. Both alternative binning and registration reduction methods typically use retrospectively acquired data, allowing their application to a large number of samples. An alternative acquisition technique must be approved by human subjects institutional review boards for patient testing; image simulations or anthropomorphic phantoms⁽⁵⁷⁾ may be used, but do not truly reflect the results obtained from patients.

I. Alternative Binning Methods

i) Amplitude Binning

4D CT post-processing techniques sort images by breathing phase to derive an estimate of the full anatomic deformation present during radiation therapy. Whether the images are reconstructed at phase definitions or sorted based on their nearest neighbor phase definition, breathing phases need to be defined accurately to minimize delineation errors.⁽⁵⁸⁾ Clinically breathing phases are ‘phase binned,’ which bases the phase definition on time relative to a breathing period; the peak inhalation points that determine the breathing period are defined by a user. Many studies have tested alternative phase definitions with the goal of increasing image accuracy while decreasing artifact presence. The most common alternative phase definition is amplitude binning, which typically relies on abdominal height or amplitude on the respiratory trace instead of time.

Rietzel and Chen⁽³⁴⁾ created software to define phases based on the percentage of displacement between user-defined peak inhalation and peak exhalation for cine 4D CT cases. They found that manual placement of peak

inhalation and peak exhalation led to improved images compared with the commercial system's automatically selected peaks, and that amplitude-sorted component phase image sets contained fewer artifacts. For proof of artifact reduction, a sample set of coronal images between methods was presented with a general statement that improvement had been observed. The same evaluation found helical amplitude binning to produce superior image quality over phase binning in a study reported by Fitzpatrick et al.⁽³³⁾ Lu et al.⁽³⁸⁾ not only found an overall image quality improvement with cine amplitude binning, but validated the greater accuracy of amplitude binning over phase binning through spirometry measurements. Wink et al.⁽⁵³⁾ found artifact improvement with amplitude binning on helical 4D CT, but only external surrogate-derived parameters were used to estimate phantom image quality. Amplitude binning accuracy has also been evaluated by measuring the consistency of a phantom sphere, with sample patient images exhibiting artifact reduction from phase binning.⁽³⁰⁾

One study⁽⁴¹⁾ compared phase binning to amplitude binning locally (as in previous studies) and globally, with a single peak inhalation and peak exhalation definition based on percentiles. An expert evaluated artifact severity in patient cases and concluded that global amplitude binning resulted in the poorest image quality and that image quality was comparable between phase binning and local amplitude binning. This study limited manual assessment to one observer who scored 4 anatomic regions; thus, this evaluation was highly subjective and could underestimate artifact frequency, but it does offer an interpretable estimate of image quality. Another comparison study of binning methods used three experts

to evaluate coronal-view image quality between percentile-based amplitude binning and the commercially available phase binning.⁽⁴⁴⁾ The experts evaluated only image areas corresponding to breathing irregularities in the respiratory trace; the methods used to identify breathing irregularities were not stated. This limitation of image regions underestimates the frequency of artifacts while relying on the assumption that externally derived respiratory trace irregularities are both directly correlated with artifact occurrence and are the only cause of artifact occurrence.

Langner et al.^(39, 46, 47) evaluated individual phase and amplitude binning methods and also evaluated combinations of those and velocity-based binning methods through 4D CT acquisition simulations. The authors incorporated a method to account for breathing irregularities and only simulated images when the respiratory signal fell within a chosen tolerance level relative to a reference signal that was derived from a least-squares fit of Fourier analysis from at least 10 patient breathing cycles. Actively avoiding the acquisition of data during breathing irregularities led to improved images over processing that lacked this avoidance, and a combination of amplitude and velocity binning yielded the best image quality among the binning techniques. Pan et al.⁽³²⁾ also used an alternative processing method that accounted for breathing irregularities retrospectively. Images associated with visually identified irregularities were flagged and were not made available for final phase binning. That report included sample images of artifact reduction achieved with this method.

ii) Correlation-Based Processing

The binning methods discussed use a respiratory trace combined with image occurrence information to yield a 4D CT with component phase image sets; no image data are incorporated into the binning process. A few studies have taken a more direct approach to artifact reduction by integrating a correlation-based evaluation of artifacts in potential image combinations before the final image sets are selected.

In one such study, Johnston et al.⁽⁵⁹⁾ sorted 10 cine 4D CT scans with amplitude and phase binning alone, as well as each binning method extended to include not only the nearest neighbor but also all neighbors within a tolerance around the nearest neighbor. This approach yields multiple potential phase image sets with the final phase sets dictated by the highest correlation coefficients; the higher the coefficient for a given image set, the less probable it is that that image set contains artifacts. Two experts visually evaluated a sample of traditionally binned images simultaneously against the correlation-sorted images and concluded that the addition of the correlation parameter reduced artifacts.

Use of the normalized cross correlation coefficient (NCC) to ‘daisy chain’ image segments has also been studied. Carnes et al.⁽²⁸⁾ acquired cine image segments with an overlapping slice to provide an anatomic link for selecting the highest NCC value among available images. These images were then ‘daisy chained’ by matching image segments one at a time beginning inferiorly, as opposed to the Johnston study, which used an algorithm to optimize the highest correlation path for the entire 3D CT. A root-mean-square difference of NCC

values was the quantitative artifact evaluation metric, and overall improvement in image quality was manually noted. A similar study evaluated image quality according to registration displacements and also found improvement in image quality.⁽⁶⁰⁾

II. Registration Methods

Registration and alternative computational methods have also been intensively studied as tools for reducing and often eliminating 4D CT artifacts in retrospectively gathered data. Hertanto et al.⁽³⁷⁾ amplitude-sorted cine 4D CT images and then registered a gap-free reference phase image set to each component phase set, extracted a motion model using principle component analysis, and applied the model to synthesize gap-free amplitude-sorted images. Another study followed this method but synthesized all phase-sorted component image sets by using a demons algorithm.⁽³⁵⁾

Alternatives to a 1D external breathing surrogate such as internal motion models⁽⁹⁾ and 3D skin models⁽⁶¹⁾ have been developed with registration to improve respiratory trace accuracy. Registration models have been applied to generate more accurate 4D CT phase sets by interpolating images at defined phases, yielding reduced artifacts in comparison to nearest neighbor phase sets.^(1, 40) Gianoli et al.⁽³⁶⁾ used a K-means clustering technique to sort cine 4D CT images with a 3D surface tracker and compared results to phase and amplitude-sorted image sets of a phantom and two patients. They concluded that the clustering technique reduced artifacts from both phase- and amplitude-sorted scans and improved proper phase

identification. Another study used NCC values, registration, and a graph-searching method to ‘stitch’ multi-slice images together to reduce artifacts.⁽⁴²⁾ Wolthaus et al.⁽⁶²⁾ registered 4D CT scans and extracted a new 3D CT of the mean of deformation positions to create an optimal primary planning dataset with reduced artifacts.

Many of these registration studies reduce, if not eliminate, artifacts by synthesizing images, but limited by the need to estimate lung CT numbers, which vary with density changes in breathing deformation,^(63, 64) limiting their usefulness of these images in treatment planning and calculations. These methods also carry a high computational cost and must have quality control in place to certify their accuracy before synthesized images can be used for treatment planning.

III. Alternative Acquisition Methods

Experimental acquisition methods constitute only a small portion of 4D CT artifact reduction studies because of the challenges associated with patient accrual and data acquisition, working with pre-existing machine settings, or both. Challenges associated with pre-existing machine settings has led to many reduction studies that use CT scanners with an extended beam width capability; challenges associated with patient acquisition may be addressed through a protocol approved by an institutional review board or by using phantoms.

Coolens et al.⁽⁴⁸⁾ acquired 4D CT scans of a phantom by using a helical 320-slice Toshiba CT scanner that covered 16 cm of superior-inferior extent and with a more conventional 16-slice Philips CT scanner. Sinusoidal and irregular patient

breathing patterns were applied and phantom volumes and centroids were assessed relative to the manufacturer's specifications. The extended extent acquisition resulted in less error than the conventional helical scan and in a faster scan time. McClelland et al.⁽⁶⁵⁾ performed a similar extended extent acquisition in combination with registration to create an accurate model of the motion states. This method acquired the entire extent in three to four image segments for five patients and compared them with the model derived with visual assessment. These investigators concluded that the model produced more accurate images, but because the extended extent acquisition was not compared with a conventional acquisition, no statement of improved quality could be made for the acquisition alone.

A series of studies using a larger cine superior-inferior extent concluded that the longer extent reduced 4D CT artifacts. A 256-multidetector row CT was used in each study that covered a 12.8 cm extent within one rotation. The first study⁽⁶⁶⁾ compared two reconstruction algorithms and their effects as evaluated within treatment plans. The second study⁽⁵¹⁾ used this acquisition method for 14 patients and evaluated image quality by contouring the GTV in each phase and evaluating the contours on fusion images with the conventional treatment planning 4D CT; GTV margin differences and linear motion extent were also evaluated between scans. That approach was found to produce a significant improvement in image quality, more accurate margins, and a shorter scanning time. A third study⁽⁵²⁾ focused on evaluating pancreatic motion in six patients and found a significantly reduced target margin and a more accurate determination of anatomic spatial

distributions. Unfortunately these studies focused mostly on target deviations and therefore underestimate the total artifact occurrence.

Low et al.⁽⁶⁷⁾ combined an experimental acquisition technique with deformable registration to generate artifact-free 4D CT scans with more accurate CT numbers. A patient with abdominal cancer was scanned on a 64-slice helical CT 25 times in alternating scanning directions. The scans were correlated with the external breathing surrogate, and a motion model was derived through a registration technique previously established by the same authors. The model was used to deform the 25 image sets into user-defined breathing phases reported to have no artifacts. Sample images of the patient scans acquired were displayed, but the model's accuracy in the generation of correct CT numbers or correct spatial distributions of anatomy is unclear.

Keall et al.⁽³¹⁾ explored the potential of artifact reduction through prospective gating by halting image acquisition for a breathing irregularity. Phantom images were acquired during a breathing irregularity, and then using beam gating during the irregularity. Visual comparison showed improved quality of a sample coronal image from the gated acquisition compared with the image acquired during the irregularity.

HYPOTHESIS & SPECIFIC AIMS

HYPOTHESIS

The 4D CT anatomic misplacement error, quantified by using a correlation-based metric, can be reduced by 30% using one of the following redundant imaging

strategies: (1) Extended acquisition, (2) Manual real-time gating, or (3) Repeat imaging of irregular segments.

Specific Aim 1: Validate an automated artifact detection algorithm. A consensus group's visual assessment of artifacts in 10 cine 4D CT scans was used as a high-quality dataset to assess the accuracy of a correlation-based artifact metric.

Specific Aim 2: Compare 4D CT artifacts in three experimental image acquisition methods. Patients received a standard clinical 4D CT scan followed by each of the three experimental methods: (1) acquiring images over two breathing cycles per couch position, (2) gating breathing irregularities out of the x-ray beam, and (3) re-scanning at couch positions associated with breathing irregularities.

Specific Aim 3: Test the reproducibility of the experimental method that reduces the artifact frequency, magnitude, or both to the greatest extent. A second cohort of patients received a clinical standard 4D CT followed by the experimental method that demonstrated the highest reduction of artifacts, as determined in Aim 2. The chosen method is repeated twice for reproducibility evaluation.

CHAPTER 2

VALIDATION OF THE ARTIFACT QUANTIFICATION METRIC

This chapter is based upon the publication,

Castillo SJ, Castillo R., Balter P, Pan T, Ibbott G, Hobbs B, Yuan Y, Guerrero T, *Assessment of a Quantitative Metric for 4D CT Artifact Evaluation by Observer Consensus* Journal of Applied Clinical Medical Physics, in press (May, 2014.)

INTRODUCTION

An artifact reduction cannot be accurately proven without a viable means of artifact quantification. To carry out Specific Aim 1, a simple, automated cine artifact metric was identified for a novel validation approach. Cui et al.⁽⁵⁶⁾ devised a correlation metric (CM) for assessment of cine artifacts that attempts to account for normal anatomical variations. This metric correlates the image slices across two neighboring couch positions and subtracts the average of correlations between adjacent image slices within the neighboring couch positions. This straightforward, efficient metric could distinguish normal anatomic variation from artifact and is thus very promising. However, validation of this metric was limited to relative metric values between 10 4D CT scans with two observers, and could be improved by a receiver operating characteristic (ROC)⁽⁶⁸⁾ analysis^(69, 70) to assess the accuracy of the metric in identifying artifacts. The current standard of artifact evaluation is manual visual assessment, which is associated with high inter-observer variability, lack of quantitative power, and lengthy analysis times. A consensus group of expert observers may reduce analysis times and inter-observer variation, yielding a higher-quality standard-evaluation dataset.

THE ARTIFACT QUANTIFICATION METRIC

For this Aim, a consensus group evaluated cine artifacts in 10 4D CT scans to form a standard dataset. This dataset was then used to assess the performance of the correlation metric (CM) in artifact identification⁽⁵⁶⁾ through ROC analysis.

I. The Normalized Cross Correlation Coefficient

The metric proposed by Cui et al.⁽⁵⁶⁾ is based on a normalized cross-correlation (*NCC*) coefficient between two axial images (Eqn. 2.1). This coefficient is commonly used in template matching to determine the position of a given pattern in an image.⁽⁷¹⁾ The position of a template \mathbf{t} within an image $\mathbf{f}(\mathbf{x}, \mathbf{y})$ is given by (\mathbf{u}, \mathbf{v}) , where $\bar{\mathbf{t}}$ is the mean of the template pixel intensities and $\bar{\mathbf{f}}_{\mathbf{u}, \mathbf{v}}$ is the mean of $\mathbf{f}(\mathbf{x}, \mathbf{y})$ pixel intensities in the region under the template. The maximum *NCC* value gives the position match.

$$NCC = \frac{\sum_{x,y} ([\mathbf{f}(\mathbf{x}, \mathbf{y}) - \bar{\mathbf{f}}_{\mathbf{u}, \mathbf{v}}][\mathbf{t}(\mathbf{x} - \mathbf{u}, \mathbf{y} - \mathbf{v}) - \bar{\mathbf{t}}])}{(\sum_{x,y} [\mathbf{f}(\mathbf{x}, \mathbf{y}) - \bar{\mathbf{f}}_{\mathbf{u}, \mathbf{v}}]^2 \sum_{x,y} [\mathbf{t}(\mathbf{x} - \mathbf{u}, \mathbf{y} - \mathbf{v}) - \bar{\mathbf{t}}]^2)^{0.5}} \quad (2.1)$$

II. The Pearson Correlation Coefficient

The Pearson correlation coefficient C is typically used in correlation studies to assess a linear relationship between two variables or images.⁽⁴³⁾ The 2D Pearson correlation coefficient was used in this study (Eqn. 2.2) for efficiency in correlating an image A with a second image B , where \bar{A} is the mean of image A pixel intensities, \bar{B} is the mean of image B pixel intensities, and (m, n) are indices of pixel

rows and columns respectively. The Pearson correlation coefficient C is equal to the maximum NCC coefficient when the two images are properly aligned. We chose to use the Pearson correlation coefficient rather than the NCC coefficient because the Pearson correlation coefficient requires less computation time and is more intuitively understood.

$$C = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \quad (2.2)$$

III. The Final Artifact Metric

All calculations were performed using MATLAB (The MathWorks, Inc., Natick, MA) with the *corr2* function (Eqn. 2) for the correlations. The correlation metric (CM) devised by Cui et al.⁽⁵⁶⁾ (Eqn. 2.3) was calculated between each couch position N per breathing phase per 4D CT scan. A couch position is a reference to a beam-width size superior-inferior location across the scan extent; each couch position contains a sorted image segment with a detector configuration of 8 x 2.5 mm thick axial images. The Pearson correlation coefficient C was calculated between image seven and image eight of couch position N ($C_{7,8}^N$), then between image one and image two of the inferior couch position $N+1$ ($C_{1,2}^{N+1}$); the resulting two coefficients were averaged to account for normal anatomical variation. The coefficient between image eight of couch position N and image one of couch position $N+1$ ($C_{8,1}^{N,N+1}$) was subtracted from this average, yielding the final metric CM . Lower CM values indicate a better image match, and a lower artifact severity.

Each 4D CT scan had a corresponding matrix of $((N_T-1) \times 10)$ CM values where N_T is the total number of couch positions per each of 10 phases.

$$CM = 0.5(C_{7,8}^N + C_{1,2}^{N+1}) - C_{8,1}^{N,N+1} \quad (2.3)$$

One metric per each couch position two through N_T-1 was needed for comparison with observer results, so each couch position's bordering CM values were averaged to mimic how an observer would assess a couch position for artifact presence. A final normalized correlation metric (NCM) was calculated by dividing each CM value by the average of CM values within that phase. This normalization allowed retention of the relative values for comparison between 4D CT scans and between phases while yielding a common reference point for 4D CT scans.

THE SAMPLE EVALUATED

I. The Sample Data

We identified 10 patients scheduled to undergo thoracic radiation therapy at MD Anderson. Our study sample consisted of the clinical cine 4D CT scans used to generate the primary dataset for each identified patient's radiation treatment plan. All 4D CT images were acquired using a GE Discovery ST PET/CT scanner (GE Medical Systems, Waukesha, WI) with the 8-slice LightSpeed CT component and were retrospectively reviewed. All image segments contained eight axial images, each with an x and y voxel size of 0.97 mm, and a z voxel size of 2.5 mm. Phases were binned evenly in time between user-defined maximum inhalation peaks for all cases. Maximum inhalation is represented by T0% with each subsequent breathing

phase per cycle defined in 10% increments of the breathing period for that cycle, i.e. T0%, T10%, T20%, etc. All 4D CT scans were phase-sorted with this binning technique by using the GE Advantage Workstation software.

II. The Sample Size Calculation

A sample size calculation estimated the minimum number of 4D CT scans necessary for metric evaluation. Two cohorts were needed, one to determine a metric threshold for artifact identification and a second to evaluate the determined artifact threshold. We calculated an exact test for single proportion (Eqn. 2.4) to determine the sample size with the software nQuery Advisor (Statistical Solutions, Boston, MA). Given x successes out of n trials, $x=0,1,\dots,n$, the p -value for an exact, two-sided binomial test under null success rate π_0 is computed by summing the probability of observing values of the sample space that are as extreme or more so than x .

$$\sum_{u=0}^x \frac{n!}{u!(n-u)!} \pi_0^u (1 - \pi_0)^{n-u} + \sum_{u=n-x-1}^n \frac{n!}{u!(n-u)!} \pi_0^u (1 - \pi_0)^{n-u} \quad (2.4)$$

A test of significance at an alpha level of 0.1 was used with type 1 error and a power of 0.81. A two-sided null hypothesis of 0.5 with an alternative of 0.75 was chosen, indicating that the metric could not definitively identify an artifact or that it could identify at least three artifacts of every four artifacts present. This yielded 26 artifacts needed per cohort. If at least one artifact exists per breathing phase per 4D CT scan, then a single 4D CT scan would contain at least 10 artifacts. Therefore, five 4D CT scans per cohort were evaluated, yielding approximately double the

number of artifacts needed. Each 4D CT scan was briefly visually assessed to ensure that this minimum artifact requirement was met.

OBSERVER ARTIFACT ASSESSMENT

I. Consensus Group

Thorough visual assessment is time-consuming and associated with high inter-observer variance,⁽⁴⁵⁾ but it remains the standard artifact evaluation method. To reduce both the time required and the inter-observer variance, we organized a committee of observers to view images simultaneously and reach a consensus on artifact location and magnitude. We termed this committee the *consensus group*. The consensus group consisted of a physician specializing in thoracic oncology, a physics assistant working with cine 4D CT, two physicists in the thoracic service, and a dosimetrist who works with thoracic treatment plans. An independent member, a graduate student studying 4D CT, coordinated each assessment and distributed materials necessary for evaluation but did not participate in the actual assessment.

II. Consensus Group Instructions

Before the first consensus group meeting, instructions for artifact evaluation were given to all of the consensus group members. The same instructions were provided at each consensus group meeting. The instructions included definitions and examples of cine 4D CT artifacts with our magnitude scoring system, as well as an example of a cardiac artifact that results from the heart beating asynchronously from the breathing motion of the lung, which could be mistaken for a 4D CT

artifact. These instructions served as a baseline reference to calibrate each observer's visual scale.

III. Artifact Evaluation Method

If a couch position within a breathing phase was identified as containing an artifact, it was assumed that the identified artifact existed at all image slices at that couch position within that phase. All couch positions covering lung anatomy were assessed; couch positions below the displayed inferior lung were considered in the assessment if there appeared to be an artifact at that location that was interfering with display of what should have been the most inferior lung anatomy. Each identified artifact was assigned a magnitude score between 1 and 5, with 1 indicating an artifact with a minor degree of interference with true anatomy and 5 indicating a large degree of interference with true anatomy; Figure 2.4 gives an example of artifact scores assigned.

Coronal views were assessed in all 10 phases for each patient. A helical deep-inspiration breath-hold scan acquired within the same examination as the 4D CT scan was displayed next to the 4D CT scan during assessment to serve as an anatomic reference, as such scans are free of the cine 4D CT artifacts being evaluated. Artifacts were assessed by using custom MATLAB software that allowed simultaneous coronal display of a component breathing phase of the 4D CT and the coronal display of the breath-hold. This software also allowed the ability to scroll through the images, change the window and level, zoom in or out, display the

couch position locations, change the breathing phase, and save an identified artifact's corresponding couch position and score (Figure 2.1).

Before the 4D CT evaluation, a lung window of level -450 HU and window 1100 HU was set for both the breath-hold scan and the 4D CT scan. Each was zoomed so that the entire lung with an extra couple of surrounding couch positions could be viewed. The couch positions of the 4D CT scan were displayed after the zoom so that observers could identify the correct couch positions. Two sets of numbered cards were distributed to each observer, and the observers used these cards to identify couch positions and scores for artifacts independent of the other observers' choices.

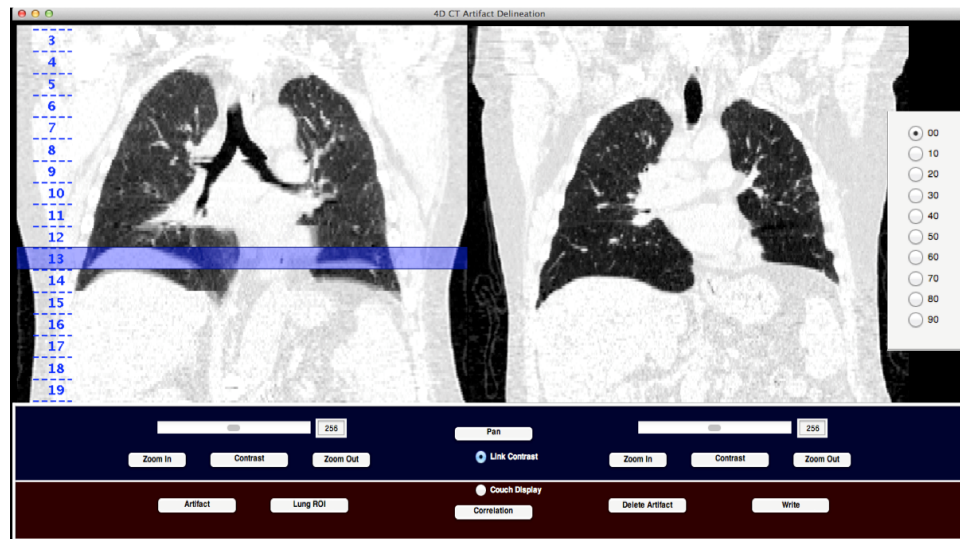


Figure 2.1. Artifact Evaluation Software. Left, T0% of a 4D CT scan with a highlighted identified artifact at couch position 13, indicating a saved artifact location. Right, corresponding deep-inspiration breath-hold scan used as an anatomic reference for artifact identification in the 4D CT (left).

The breath-hold scan was scrolled through before each 4D CT evaluation for anatomic reference, and then the entire lung in the selected 4D CT breathing phase was scrolled through, allowing time for observers to choose their numbered cards for artifact location and score. Then for each couch position within the lung, the consensus group was asked if the stated couch position contained an artifact. Each observer shared the numbered cards he or she had chosen. Results were saved if all consensus members agreed on an artifact location and score; if different answers were chosen, the observers made a case for their reasoning and the images were reviewed until the group achieved a consensus. Only rarely was no consensus reached; in those cases, the majority ruled.

IV. Analysis of Consensus-Chosen Artifacts

A logistic mixed effects model was used to evaluate the relationship between artifact incidence and breathing phase from consensus results from all 10 patient cases by using the lme4 package in R.2.14.0 (<http://www.r-project.org/>). This utilized mixed logistic model facilitates joint estimation of the log-odds of artifact incidence across all phases while accounting for the inherent correlation among observations derived from the same patient scan to estimate and characterize the extent of variability associated with a breathing phase. The likelihood ratio test for association is reported; this test is used to weigh the observed evidence of homogeneity, i.e. the identical log-odds for all phases.

In addition, phase effects were evaluated for significance using two-sided Wald tests. Wald tests are used to weigh the observed evidence of phase specific log-odds adjustments from component phase T0% in the presence of the estimated

within and between subject sources of variability. Post-hoc application of the sequentially rejective Bonferroni method was used to adjust p-values for multiple comparisons of the adjustments of log-odds of artifact incidence from component phase T0% across phases T10% through T90%.⁽⁷²⁾ The multiple phase comparisons inflate the nominal false positive rate, and thus the p-values need to be adjusted to maintain the type I error rate at 5%.

ARTIFACT THRESHOLD DERIVATION

I. ROC Curve Generation

An ROC method was used to determine an artifact threshold with cohort 1; this artifact threshold was then applied to cohort 2 to evaluate the resulting sensitivity and specificity. The artifact threshold is an NCM value. A 4D CT contains NCM values for each component breathing phase at each couch position except for the first and last as given in the matrix $((N_T-2) \times 10)$. The NCM values above or equal to the applied artifact threshold value indicate an artifact at an anatomic location (N) and breathing phase. MATLAB software was developed to format observer results and produce ROC curves.

To generate ROC curves, a binary decision threshold is moved across the data, above which an artifact is identified, while below which an artifact is not identified; each decision threshold yields a sensitivity and false-positive fraction point on the curve.⁽²⁰⁾ The decision threshold was incremented by 1% between the minimum NCM value and the maximum NCM value contained within each cohort 1 patient $((N_T-2) \times 10)$ matrix. As the decision threshold was incremented, NCM values

below the decision threshold were set to zero, indicating no artifact presence. The consensus results served as the “ground truth” to determine the true-positive fraction, true-negative fraction, false-positive fraction, and false-negative fraction.

II. ROC Curve Parameters

Parameters were calculated from the ROC curves to assess the accuracy of the NCM and find the resultant artifact threshold. For each patient in cohort 1, the area under the curve (AUC) was calculated to determine the NCM accuracy, and the Youden index was calculated to extract a threshold. The Youden index J is the maximum vertical distance between the ROC curve and the diagonal “chance line,” which can also be related back to a decision threshold point directly (Eqn. 2.5).

$$J = \max (sensitivity(dc) + specificity(dc) - 1) \quad (2.5)$$

The Youden index represents the optimal cut-point in ROC curve analysis and is used as another measure of accuracy. Youden index values vary between 0 and 1, with 1 indicating a relatively large NCM evaluation accuracy.⁽⁷³⁾ The artifact threshold was derived from the Youden index to provide the optimal artifact threshold corresponding to maximum accuracy in each ROC curve. The point in the curve at which the Youden index was found yielded the corresponding NCM threshold.

III. Artifact Threshold

The artifact threshold corresponding to the minimum Youden index found in cohort 1 was taken as the final artifact threshold. One outlier index was found, and

thus an average of thresholds was deemed inappropriate, and the minimum was taken to ensure that artifacts would not be missed.

ARTIFACT THRESHOLD EVALUATION

The determined artifact threshold was applied to each $((N_T-2) \times 10)$ matrix of NCM values in cohort 2. All cohort 1 ROC curves and parameters, and cohort 2 sensitivity and specificity values were calculated using consensus group results as the ground truth.

METRIC VS CONSENSUS-CHOSEN ARTIFACT MAGNITUDES

To evaluate whether the relative NCM values accurately reflected the degree of artifact severity, or artifact magnitude, the artifact score for each consensus group-identified artifact among all 10 cases was compared to the NCM value corresponding to the couch position and breathing phase of that artifact score. A Pearson correlation coefficient and a p-value were calculated for each patient.

BREATHING IRREGULARITY

I. Breathing Irregularity Identification

Breathing irregularity identification for this research was performed by using a simple method based on finding outliers across the scan extent. Because each couch position contains at least one breathing cycle with 10 phases, multiple breathing phases per component phase exist across the scan extent. The corresponding amplitudes of the multiple breathing phases were averaged per component phase to yield a baseline for regular breathing identification. A phase was considered

irregular when the amplitude existed outside a tolerance range set by the mean amplitude ± 1 SD of said phase (Figure 2.2).

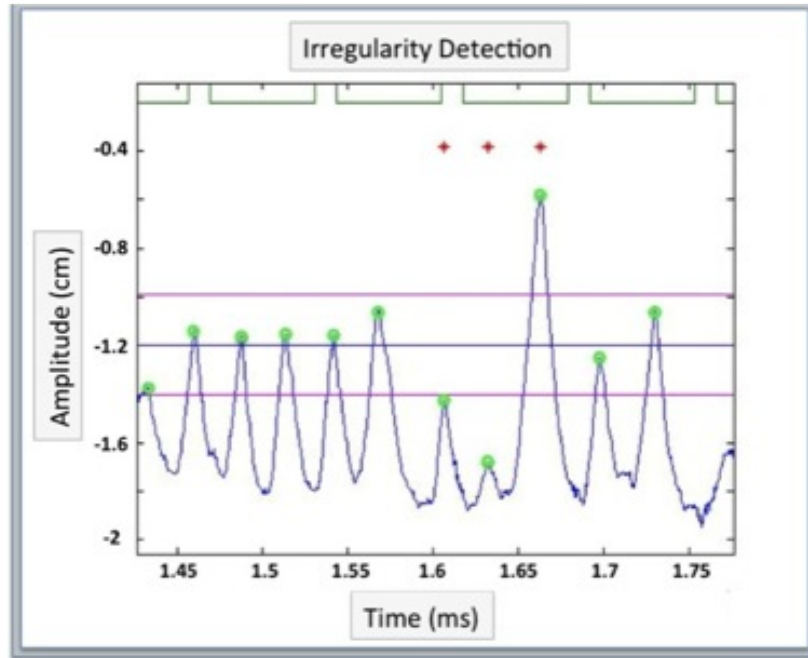


Figure. 2.2. Breathing Irregularity Identification. Patient respiration (blue), beam-on signal (green lines), T0% phases (green circles). T0% mean (middle horizontal dark blue line), T0% SD (outer pink horizontal lines), T0% irregular phases (red stars).

II. Breathing Irregularity vs Consensus-Chosen Artifacts

Three questions concerning breathing irregularities and artifacts were addressed: (1) if artifact presence is indicative of breathing irregularity presence, (2) if breathing irregularity presence is indicative of artifact presence, and (3) how artifact magnitude relates to breathing irregularities.

To answer the first question, whether or not an observed artifact corresponds to a breathing irregularity, each chosen artifact location per patient was identified and flagged if a breathing irregularity occurred during image acquisition at that location. The percentage of the total number of irregular phases occurring at a consensus-chosen artifact position per all consensus-chosen artifact positions was examined and averaged across the 10 consensus patient results.

The second question was addressed similarly to the first; the total number of breathing irregularities that occurred during image acquisition of the lungs was flagged, as well as the artifacts identified corresponding to those irregularity locations. The percentage of the total number of consensus-chosen artifacts that occurred at breathing phase locations identified as irregular per total number of irregular breathing phase locations was calculated. The distribution of mean artifact and irregularity quantities per phase per patient was also evaluated.

Finally, to compare artifact magnitudes, or scores, against irregularity presence we tallied the number of artifacts per score across phases per patient associated with a breathing irregularity and unassociated with a breathing irregularity. The 10 values per score were averaged and correlated with score.

RESULTS

I. Artifact Metric

Evaluated cases were labeled first by the cohort (C1 or C2) followed by a number indicating the order of evaluation within that cohort. An example of NCM values over all couch positions for two component phases for the first patient of cohort 2, C2_1, is shown in Figure 2.3; coronal image slices from the same two

component phases of the C2_1 case are shown in Figure 2.4 for comparison to Figure 2.3 and as an example of consensus-chosen scores.

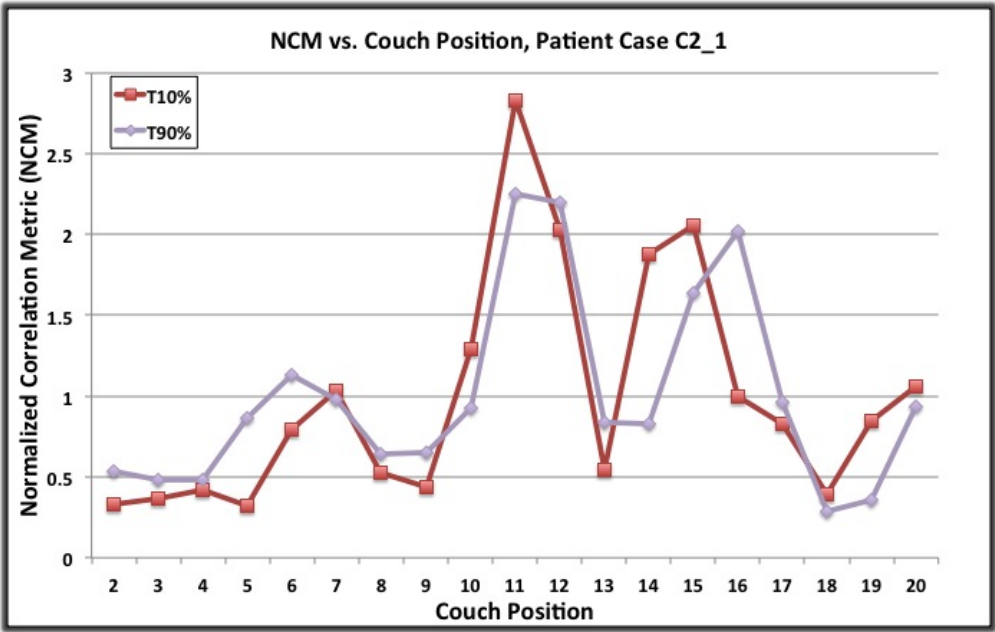


Figure 2.3. Normalized Correlation Metric vs Couch Positions, Patient C2_1.

Breathing phases T10% and T90% are displayed for comparison with coronal views of T10% and T90% in Figure 2.4.

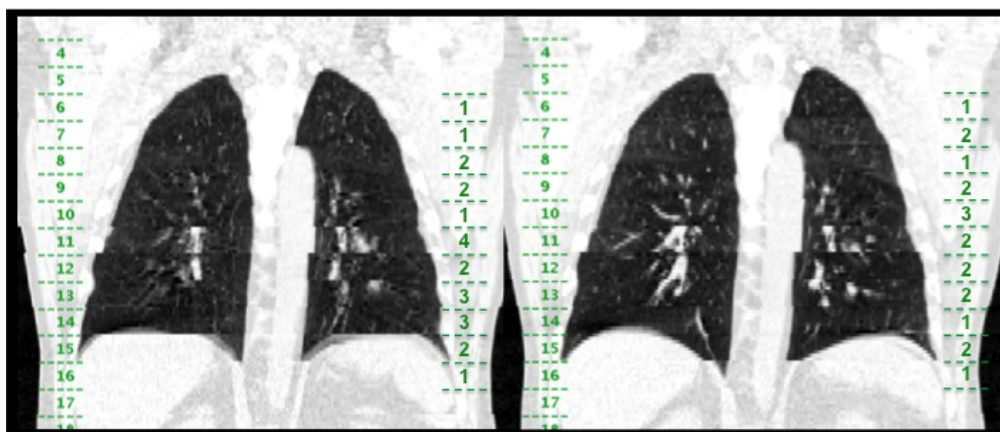


Figure 2.4. Example Coronal Slices of Patient C2_1. Examples are T10% (left)

and T90% (right) (NCM values shown in Figure. 2.3), with couch positions

indicated on the left side of each coronal view and consensus group identified

scores per couch position shown on the right side of each coronal view. The artifact

in T10% at couch position 11 was scored as being more severe based on the higher

interference of the artifact with anatomy.

II. The Consensus-Chosen Artifacts

The consensus group scored all 10 phases of a 4D CT scan in less than 40 minutes, on average within 30 minutes. The mean percentage of couch positions covering lung anatomy that contained consensus-chosen artifacts for both cohorts was 68.7%. The first patient scored by the consensus group had the lowest percentage of lung couch positions containing artifacts at 32.4%; all other patients had artifacts in at least 59% of lung couch positions with a maximum percentage of 87.3%.

Artifact incidence was significantly associated with breathing phase by using logistic mixed effects analysis ($p < 0.002$ likelihood ratio test). The estimated

probability of observing an artifact during the T60% phase was significantly lower than T10% (odds ratio=0.44, $p<0.003$), T70% (odds ratio=0.41, $p<0.0015$), T80% (odds ratio=0.33, $p<0.0002$), and T90% (odds ratio=0.36, $p<0.0004$) phases after adjusting for multiplicity. Figure 2.5 demonstrates the model-estimated probability of an artifact at each phase with 95% confidence intervals.

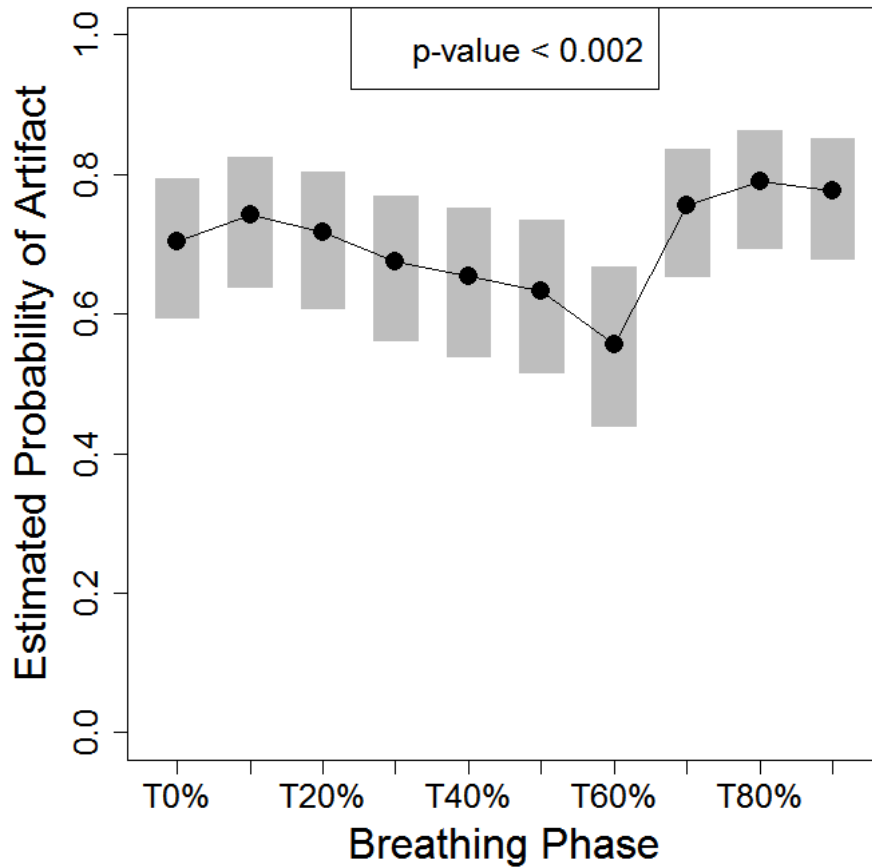


Figure 2.5. Estimated Probability of Artifacts as a Function of Phase.

Associated 95% confidence intervals indicated by grey bars. The risk of an artifact was lower for exhalation phase images. The p-value derives from the likelihood ratio test of the global null hypothesis of the absence of an association with phase.

III. ROC Analysis

Figure 2.6 demonstrates a sample ROC curve generated for case C1_5, with the associated parameters; Table 2.1 displays the ROC curve parameters for all five patients in cohort 1. The artifact threshold found using this method was 73%. The average (AVG) cohort 2 sensitivity resulting from the applied artifact threshold of 73% was 0.703, and the average specificity of cohort 2 was 0.476, each with a standard deviation of 0.11. All cohort 2 values are given in Table 2.2.

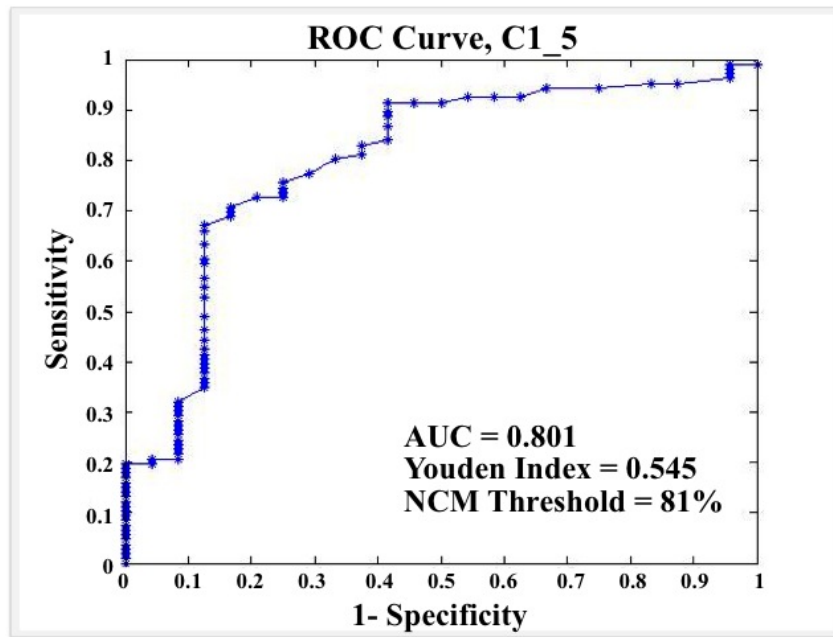


Figure 2.6. Example ROC Curve from C1_5. Parameters displayed are area under the curve (AUC), Youden index, and corresponding NCM threshold.

Parameter	C1_1	C1_2	C1_3	C1_4	C1_5
AUC (Area Under Curve)	0.756	0.525	0.769	0.709	0.801
Youden Index	0.446	0.103	0.507	0.461	0.545
NCM threshold	125%	73%	93%	93%	81%

2

Table 2.1. Cohort 1 ROC Parameters

	C2_1	C2_2	C2_3	C2_4	C2_5	AVG ¹	SD ²
Sensitivity	0.602	0.840	0.777	0.707	0.589	0.703	0.109
Specificity	0.515	0.630	0.388	0.357	0.490	0.476	0.109

Table 2.2. Cohort 2 Sensitivity and Specificity Values

¹ AVG: average, ² SD: standard deviation.

IV. Metric vs Consensus-Chosen Artifact Magnitudes

The mean of cohort 1 Pearson correlation coefficients was 0.80, with all but the first case yielding a coefficient greater than 0.91; the composite correlation coefficient was 0.54. The mean of cohort 2 Pearson correlation coefficients was 0.61, with three of the five cases yielding coefficients greater than 0.99; the composite correlation coefficient was 0.58. The composite p-values for both cohorts were less than 0.001, indicating a nonzero correlation between the NCM and consensus scores.

V. Breathing Irregularities vs. Consensus-Chosen Artifacts

The mean percentage breathing irregularities occurring at the same couch location and phase as each consensus-chosen artifacts across all patients was 37.8% (range: 23.7%-53.5%); values per phase are shown in Figure 2.7. The mean percentage of consensus-chosen artifacts occurring at the same couch location and phase as each breathing irregularity across all patients was 68.4% (range: 49.7%-81.6%); values per phase are shown in Figure 2.8. The total numbers of consensus-chosen artifact scores (range 1-5), with and without corresponding breathing irregularities, are shown in Figure 2.9.

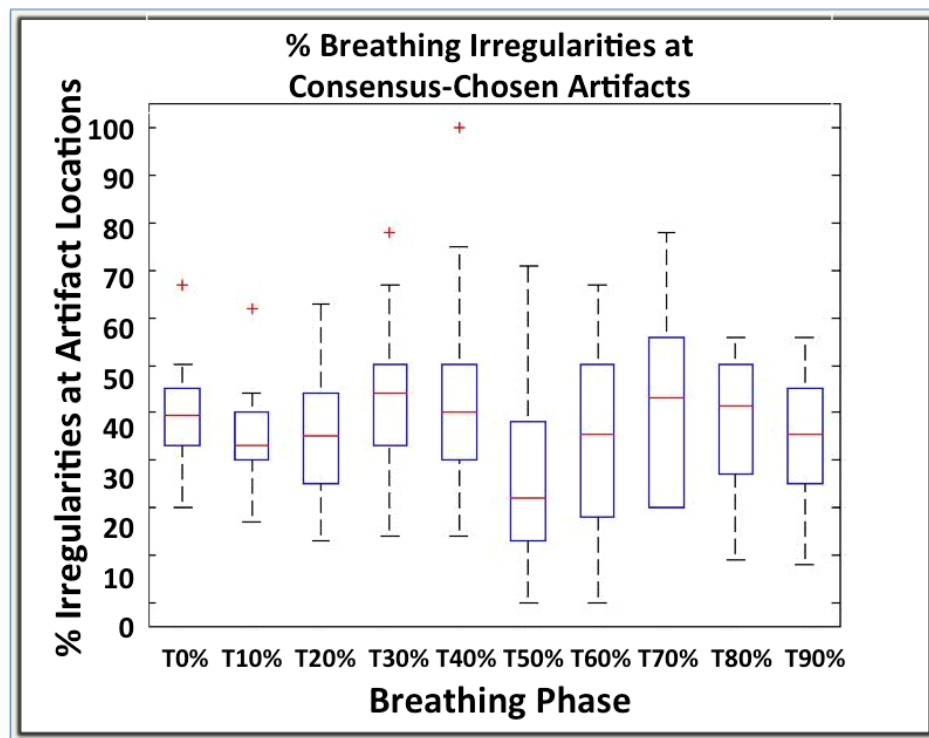


Figure 2.7. Percentage Breathing Irregularities at Artifact Locations. Boxplot of percentage of breathing irregularities that occur at the same location as consensus-chosen artifacts.

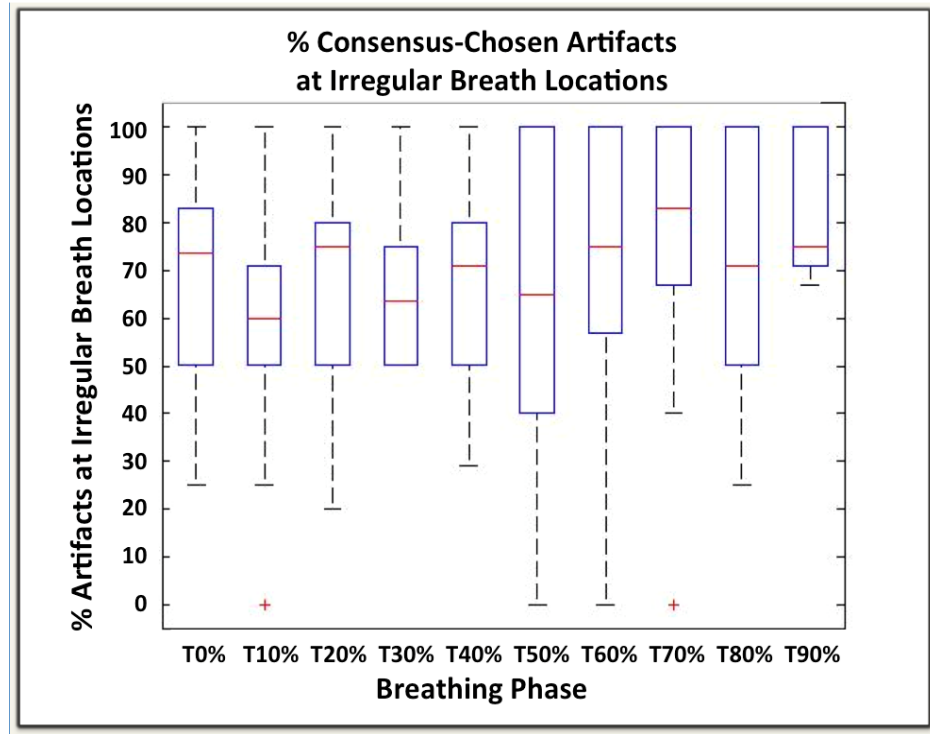


Figure 2.8. Percentage Artifacts at Irregular Breath Locations. Boxplot of percentage of consensus-chosen artifacts that occur at the same location as breathing irregularities.

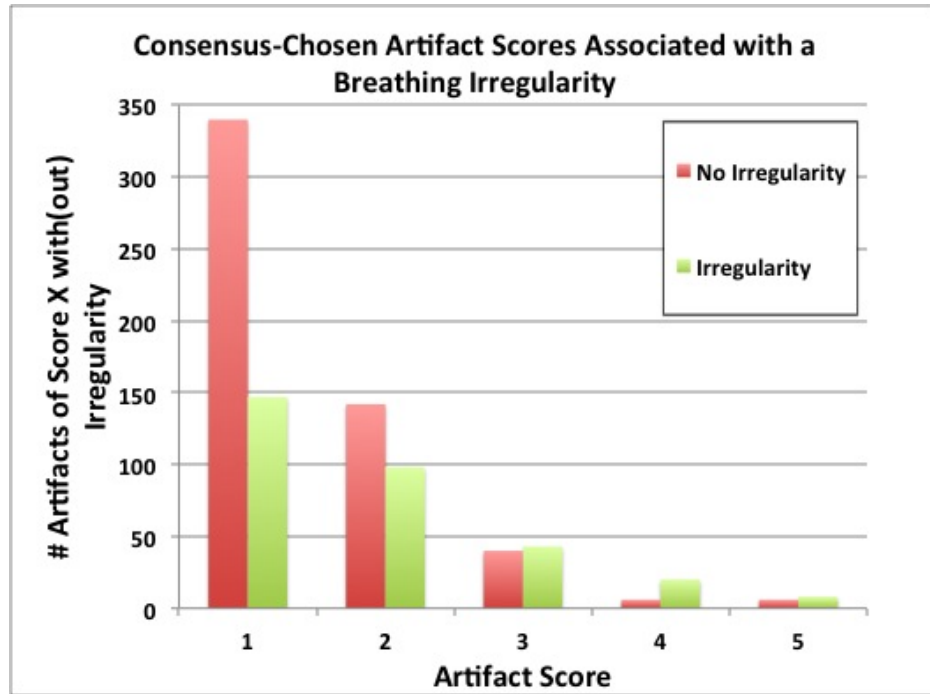


Figure 2.9. Artifact Scores and Breathing Irregularities. Total number of consensus-chosen artifact scores with and without a corresponding breathing irregularity.

DISCUSSION

We found that the average sensitivity and artifact score correlation achieved for the NCM assessment against the consensus group assessment indicated that the NCM method performed moderately well for evaluation of cine 4D CT artifacts. The sensitivity was high but the specificity was only moderate, suggesting that the artifact threshold found using cohort 1 overestimated the true number of artifact locations. This was expected, because we chose the artifact threshold corresponding to the minimum Youden index, the cohort 1 outlier, to favor not missing an artifact at the expense of falsely identifying a position as containing an artifact. However, even with this trade-off, the sensitivity was not as high as expected. This may be due to the inexperience of the

consensus group members present with the scoring system for the first few cases. If the consensus group members had participated in a practice session an outlier might have been avoided, and the artifact threshold would have been derived over an average of the sample rather than from the outlier value. Had a larger number of patients in cohort 1 been evaluated, a decreased outlier detriment would have allowed an average-derived artifact threshold.

The artifact score correlations between the NCM values and consensus results were good on an individual basis, but a few cases with poor coefficients reduced the composite correlation coefficient. The first case had a poor coefficient (0.169) that, as noted above, would have been improved if a practice session had been included to familiarize the observers with the procedure and scoring system based on the given instructions. The other two cases with poor coefficients (0.342 and 0.408) had no consensus magnitude scores higher than 2, which left only two data points for the correlation. An increase in number of patients in cohort 1 might also have provided more cases with scores over the range 1-5 to offset these poor correlations.

This study would have benefited from a practice session to train observers before the consensus evaluations and an increased cohort 1 sample size of patient 4D CT cases to offset the effect of an outlier and provide more data points for artifact score correlation. However, even with these weaknesses, the consensus group method of visual assessment allowed an efficient and guided scoring that produced a high-quality research dataset for metric validation, reduced inter-observer variation, and provided a more consistent method of identifying artifact locations and magnitudes. The NCM also performed well despite these issues and resulted in a reproducible automated quantitative evaluation

within a time that could not be achieved by observer evaluations; the NCM is also simple to implement as it is based on the Pearson correlation coefficient, and thus would integrate well into the clinical workflow.

The results of the logistic mixed effects analysis demonstrated that artifacts were less likely to be present near exhalation. This is intuitive, because exhalation is a more stable breathing state; passive exhale ends at the functional residual capacity, the equilibrium point between the chest wall expansion and lung contraction. In particular, a significant reduction in artifact odds was noted at T60% when compared to T10%, T70%, T80%, and T90%. These images were phase-sorted using 10 equally divided bins, and did not explicitly define T50% to be maximum exhalation. As time from peak inhalation to peak exhalation is typically longer than time from peak exhalation to peak inhalation,⁽⁷⁴⁾ T60% may have more truly represented maximum exhalation than other phases such as T50%,⁽⁷⁵⁾ providing an explanation for why only T60% contained a significant artifact reduction.

The relationship between breathing irregularities and artifacts remains, but is not exact. The presence of a breathing irregularity indicates the presence of an artifact in roughly two-thirds of cases, whereas the presence of an artifact was associated with an irregularity roughly a third of the time. An irregularity causing an artifact to occur has a strong association, but irregularities may not always be the cause of an artifact. This is intuitive because artifacts are not always caused by a breathing irregularity; an artifact may also be caused by the lack of RPM correlation, phase mis-assignment errors, or from the acquisition or post-processing technique. We expected that higher-magnitude artifacts would be associated with a higher number of breathing irregularities, but this was only

apparent for the score of 4. This may reflect the lack of data points for the higher scores, particularly for score 5, or the lack of a more robust irregularity detection that would also allow a magnitude to be scored.

Currently, visual assessment is the evaluation standard, but it lacks a clear set of rules, which makes extraction of quantitative results difficult. Some reports of visual assessments display sample images with a description of the overall artifact presence whether on an individual basis or as a comparison between scans,^(1, 28-32, 34, 36-40) whereas others state a particular method of an independent expert evaluation.^(26, 41-44) Our study extracted a quantitative artifact evaluation from qualitative observations in a guided and efficient method that reduced inter-observer variability. This yielded a high-quality research dataset for 10 breathing phases of 10 patients in approximately five hours. This consensus group method of evaluation could also be used for images from various acquisition and processing methods and even for various types of artifacts, provided that instructions on identification and characterization are provided, preferably with a practice session before the first case analysis.

Similar evaluations have been done with images derived from phantoms and software simulations, in which the exact anatomical volume and shape are known.^(30, 35-37, 46-48) Persson et al. evaluated artifacts in terms of GTV deviations in comparison with a reference target volume.^(49, 50) Other quantitative artifact evaluations include: deviations in target centroid position or contours,^(39, 51, 52) the mean square gray value difference between couch positions,⁽¹⁾ tidal volume variations from reference images,⁽³⁸⁾ and external surrogate parameters.⁽⁵³⁾ The quantitative metric evaluated in this study offers an efficient, reliable method to evaluate cine artifact location and magnitude in the lung.

CONCLUSION

We conclude that the correlation metric assessed has promise for use in the evaluation of artifacts on cine 4D CT scans, when an efficient and reliable method is needed for processing many sets of images, although additional cases would yield an even more accurate artifact threshold for identification. We also conclude that the consensus group method has the potential to be used as a research standard for evaluating 4D CT artifacts and as a standard to evaluate alternative quantitative artifact evaluation methods, and we recommend that the consensus group have a practice session before evaluation.

CHAPTER 3

EXPERIMENTAL 4D CT ACQUISITION METHODS

INTRODUCTION

4D CT artifacts cause uncertainty in the true anatomic spatial distribution and could lead to errors in treatment planning delineation and targeting; these artifacts have been demonstrated to affect emerging applications such as lung function images derived from CT.^(26, 27, 49, 50, 76, 77) In addition to artifacts common in diagnostic CT, 4D CT images are subject to artifacts caused by irregular breathing^(1, 26, 27, 31, 32, 37, 49, 58); irregular breathing is currently not corrected for in the clinical setting. Although breathing irregularities may introduce appreciable artifacts into the 4D CT, the current clinical strategy is to use the low-quality dataset for treatment planning or to re-acquire the 4D CT scan. If the re-acquired 4D CT still demonstrates appreciable artifacts, free-breathing helical acquired CT images are used to aid in treatment planning. The use of these images was standard practice prior to the introduction of 4D CT methods, but it does not account for respiratory motion.

Several research groups have reported methods to reduce 4D CT artifacts by reducing the data that are acquired during breathing irregularities. Langner et al.^(39, 46, 47) compared sorting methods to reduce respiratory motion artifacts in a retrospective simulation study. A model simulated CT images when a patient respiratory signal fell within a tolerance limit of a reference respiratory trace, and only those images were used in the sorting process. By simulating images only when the respiratory waveform was within the tolerance limit of the reference signal, a higher quality 4D CT was produced compared to retrospective phase sorting using all of the images.^(39, 46, 47)

Pan et al.⁽³²⁾ reduced cine 4D CT artifacts by identifying the data acquired during irregular breathing and disabling the use of that identified data in phase sorting. Sample coronal and sagittal views visually demonstrated improved image quality when these irregular portions were excluded; however this was only applied to one region of the scan extent.

Keall et al.⁽³¹⁾ explored the potential for artifact reduction through prospective gating by halting the image acquisition for a breathing irregularity. Coronal images of a phantom were acquired both during a breathing irregularity and by using beam gating to exclude the irregularity. The image acquired with gating was found by visual comparison to be of better quality than the image that was not gated.

Despite these promising results, to the best of our knowledge, methods to prospectively reduce 4D CT artifacts by altering the acquisition have not been attempted in a clinical setting. In the present study we implemented three experimental 4D CT acquisition methods that target breathing irregularities: (1) acquiring more images, (2) gating the x-ray beam with breathing irregularities, and (3) re-acquiring images associated with breathing irregularities in patients with thoracic cancer to determine the potential for improving 4D CT image quality in a clinical setting in a relatively simple manner. Unlike prior studies, the methods used in this study focused on acquisition modification rather than retrospective analysis.

I. The Patient Sample

With the approval of The University of Texas MD Anderson Cancer Center institutional review board (protocol 2011-0631), the first phase of this study

involved 18 patients scheduled to receive thoracic radiation therapy. The patients (eight women and 10 men) had clinical diagnoses of non-small cell lung cancer (N=12), esophageal cancer (N=4), or mesothelioma (N=2). The mean (\pm standard deviation [SD]) age of study participants was 66.3 ± 10.1 years. Each patient received a standard simulation 4D CT scan, immediately followed by each of the three experimental acquisition methods.

II. 4D CT Parameters

All 4D CT images were acquired in cine mode on a GE Discovery ST PET/CT scanner (GE Medical Systems, Waukesha, WI) with a 500 ms tube rotation time; the CT component is an 8-slice Lightspeed CT. A real-time processing monitor (RPM; Varian Associates, Palo Alto, CA) served as an external surrogate for organ motion. This provided a respiratory trace of relative abdominal height versus time. The acquisition time per couch position (cine duration) was based on the patient's average breathing cycle plus 1 s. This acquisition over the patient's breathing cycle yields several multi-slice image segments (8 \times 2.5 mm axial images) per couch position. Images were reconstructed by using 360° of data (full reconstruction) with a cine time between images ranging from 350 ms to 450 ms. Scans were obtained at 120 kVp, with a 100 mA tube current for the clinical acquisitions and a dose-sparing 50 mA for the experimental acquisitions. The tube current reduction for the experimental acquisitions was deemed appropriate due to the large increase in the total imaging dose from the additional scans and the minimal effect that the reduction in the beam current would have on the image quality.⁽⁷⁸⁾

III. Experimental Methods

i) Gating Method

To implement the gating method (GM), we monitored the respiratory trace in real time to facilitate manual beam gating when an apparent breathing irregularity occurred, and we subsequently restarted the beam when regular breathing returned. Manually stopping the beam caused the entire couch position to repeat to ensure that at least one breathing cycle was acquired per couch position; this yielded a full 4D CT image set plus gated sections.

ii) Rescan Method

For the rescan method (RM), image segments were re-acquired at identified locations at which breathing irregularities occurred during the clinical acquisition. Implementation of the RM involved immediate post-processing of the clinical 4D CT respiratory trace to identify the couch positions where images were to be re-acquired. Only the identified couch positions were rescanned and replaced in the clinical 4D CT for the final RM image set.

iii) Oversampling Method

For the oversampling method (OM), the beam-on time was increased to capture two breathing cycles per couch position, or approximately double the data of the clinical scan; therefore the clinical cine duration was increased by a factor

of two unless the patient's breathing had changed. This was implemented by altering the manufacturer's internal constraint for data acquisition from 3000 to 6000 images to allow for extended cine durations.

iv) Breathing Irregularity Detection

To identify breathing irregularities, we defined a tolerance range that was based on the mean and standard deviation of the 10 phase amplitudes across the scan extent. A phase was considered irregular when the amplitude fell out of the range given by mean amplitude ± 1 standard deviation. A couch position warranted repeat imaging if 30% or more of the location was associated with irregularities. This threshold was chosen based on an average of 15 image segments per couch position; if 30% or more of the segments are irregular, then the final phase image set will contain segments associated with an irregularity that is likely to be present as an artifact. An example of an irregular T0% phase in a patient respiratory trace is shown in Figure 2.2. Breathing irregularities were calculated retrospectively in all 4D CT acquisitions. This method of irregularity detection is been outlined in Chapter 2 section I.

v) Acquisition Order

The clinical acquisition was always done before the experimental scans; because the RM combined data from an independent scan into the clinical scan, this method was always performed immediately after the clinical scan to minimize any changes in breathing or internal and external anatomic shifts. Either

the OM or GM followed the RM; these were alternated in an attempt to offset any detrimental effect from the end of the exam acquisition. A normal inhalation breath-hold was taken as an inhalation breath-hold as close to the patient's normal T0% maximum inhalation as possible. A deep inspiration breath-hold is the clinical standard, but a deep inspiration breath-hold exaggerates the inhaled anatomic state. This yields a larger vector magnitude between identical anatomy from the normal inhalation T0% and from the deep inspiration breath-hold; the normal inhalation breath-hold yields a smaller vector magnitude from the T0%.^(79, 80)

IV. Post-Processing Methods

ii) Phase Sorting

Each acquisition method was processed in two ways: by phase sorting and by a quantitative image correlation-based sorting method. The T0% phases were defined by selecting the maximum inhalation peaks on the RPM respiratory trace; the subsequent breathing phases were defined as percentages of the breathing period in 10% increments, i.e. T0%, T10%, T20%, etc. The clinical standard sorting method is phase sorting, as outlined in Chapter 1. Custom MATLAB (The MathWorks, Inc., Natick, MA) software was created for phase sorting of clinical and experimental acquisitions. The GE Advantage Workstation is used for phase sorting 4D CT scans that will be used in treatment planning. We processed our clinical acquisitions by using our custom software to ensure that the protocol clinical acquisitions and experimental acquisitions could be compared accurately

without differences from software implementations adversely affecting the results.

A few minor differences in phase sorting exist between our software and the GE software: the GE software modifies the phase definitions to occur directly at RPM samples (the RPM sampling period is approximately 33.4 ms) and the first of two image segments is chosen when the two segments occur an equal time distance from a defined phase. The former difference may result in one or two of GE-generated couch positions containing different image segments than our custom sorted scans because of a potential difference in the phase definition of up to 17 ms. Retention of the exact percentage of each breathing cycle period for phase definitions results in more accurate image sorting. The latter difference typically only would affect the image segment chosen for one couch position, but this situation is very rare and does not occur for many patients or breathing phases.

ii) Experimental Sorting

The experimental sorting method is an expansion of phase sorting: instead of choosing the nearest neighbor image segment to a phase, multiple image segments are chosen and an optimal combination is determined by minimization of the artifact metric (Eqn. 3), as validated in Chapter 2. The breathing phases are defined the same as in phase sorting. The three closest image segments in time to the phase definitions were binned; this allowed retention of breathing information with potential for artifact reduction in the available data. The absolute value of the

sum of the correlation metric (CM) values across all of the couch positions was minimized by using the shortest path Dijkstra's algorithm.^(30, 31) This sorting method is referred to as CM sorting. Custom MATLAB (The MathWorks, Inc., Natick, MA) software was developed for this experimental sorting method.

V. Artifact Evaluation

i) Artifact Analysis

The CM provided quantitative-based rankings of scan methods per patient per sorting technique. The sum of the absolute value of CM values was calculated for each component phase image set of each acquisition method for each sorting method, and rankings were derived from the mean of the summed CM values over the phases. A rank (1-4) was assigned to each method per patient, with increasing rank corresponding to increasing averaged CM value and poorer image quality. The images were visually assessed to verify the rankings.

ii) Statistical Analysis

Statistical analysis of the average CM values used one-way mixed ANOVA with heteroscedastic variance by acquisition method. An F-test was used to test for association between acquisition methods. Pairwise comparisons among the four acquisition methods used simultaneous inference with Tukey's method. The familywise error rate was controlled at the 0.05 significance level. Tukey's adjusted, a two-sided p-values are reported. Interval estimation of the percentage reduction in CM values for CM sorting relative to phase sorting averaged over

acquisition method and phase is provided with two-sided 95% confidence intervals.

VI. Effective Dose Estimates

The estimated effective dose delivered to each patient using each of the 4 acquisition methods was also calculated, adjusting for clinical tube current and clinical acquisition extent in the experimental methods for accurate comparison with the standard clinical acquisition. Dose estimates were derived from the CTDI method^(81, 82) based on a measured CTDI_{vol} of 50 mGy⁽³²⁾ for a cine duration of 5.6 seconds, and a tube current of 100 mA with a k-factor for an adult chest of 0.014 mSv/mGy-cm^(81, 83, 84).

VII. Artifact Presence vs Breathing Irregularities

The presence of artifacts as indicated by mean CM values over the phases was correlated with the number and percentage of irregular breathing phases over the scan extent to determine whether an overall poor image quality corresponded to overall poor breathing regularity in clinical and oversampling phase-sorted and CM-sorted acquisitions. The number of breathing irregularities was also correlated with the percentage of GM acquisition duration compared with the clinical acquisition, to determine whether longer scan times were required because of highly irregular breathing. Pearson correlation coefficients were used in these calculations.

RESULTS

I. Patient Summary Statistics

The mean percentage of breathing irregularities present during image acquisition for all 72 4D CT scans was $28.0 \pm 7.7\%$. Respiratory trace parameters derived over the scan extent included a mean displacement from inhalation to exhalation of 0.98 ± 0.41 cm, and an average breathing period of 4.34 ± 1.3 seconds. The mean percentage of the re-acquired scan extent for the RM was $40.3 \pm 10.0\%$; a mean of $37.8 \pm 18.1\%$ of the images were re-acquired in the lung. A mean of $28.4 \pm 11.2\%$ of breathing irregularities occurred during the re-acquisition of images for the RM. A mean of $54.6 \pm 12.1\%$ of breathing irregularities were not gated during the GM image acquisition, which is a 26.7% reduction from the clinical acquired irregularities. The GM led to a $66.7 \pm 39.2\%$ increase in acquisition time compared with the clinical 4D CT acquisition time; this was due to the couch repetition for each beam stop and the time intervals between x-ray beam on during irregular breathing. Automatic x-ray beam control and acquisition of only one cine duration per couch position would decrease this prolonged extension of the scan duration. Despite the increase in scan duration, all clinical and experimental 4D CT scans were acquired within the standard clinical 60-minute time slot.

II. Experimental Acquisition Rankings

i) Phase-Sorted Acquisitions

Among phase-sorted acquisition methods (Figure 3.1) the GM and the clinical ranked highest (each $N=8$), followed by the OM ($N=2$).

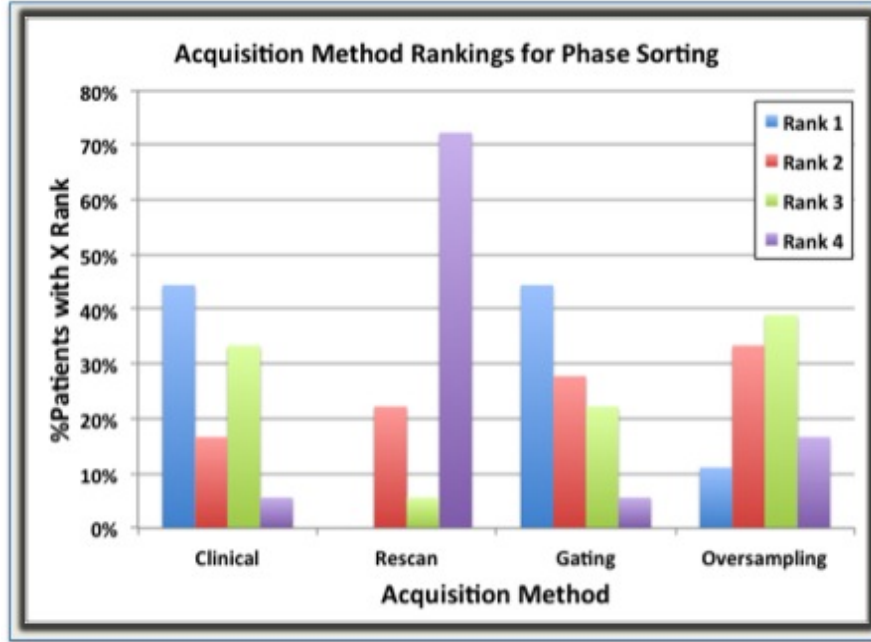


Figure 3.1. Phase-Sorted Acquisition Rankings. Distribution of rankings for the phase-sorted 4D CT methods for the 18 patients. Rank=1, blue bars, 1st from left; Rank=2, red bars, 2nd from left; Rank=3, green bars, 3rd from left; Rank=4, purple bars, 4th from left

ii) CM-Sorted Acquisitions

Among CM-sorted acquisition methods (Figure 3.2) the OM ranked highest (N=9), followed by the GM (N=5), and the clinical (N=4). Sample coronal images comparing the CM-sorted acquisition methods (Figure 3.3) and comparing the sorting methods (Figure 3.4) are provided.

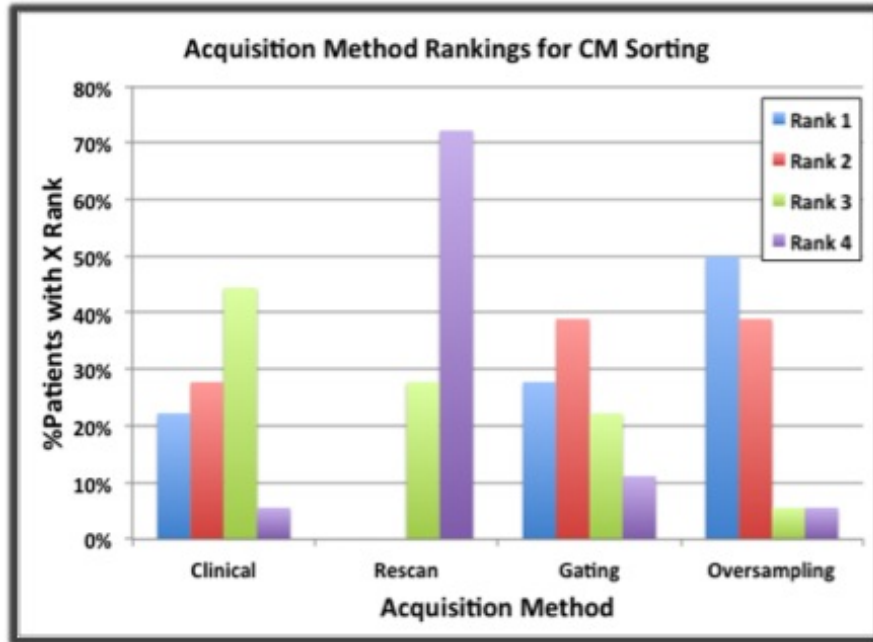


Figure 3.2. CM-Sorted Acquisition Rankings. Distribution of rankings for the CM-sorted 4D CT methods for the 18 patients.



Figure 3.3. CM-Sorted Sample Coronal Views.

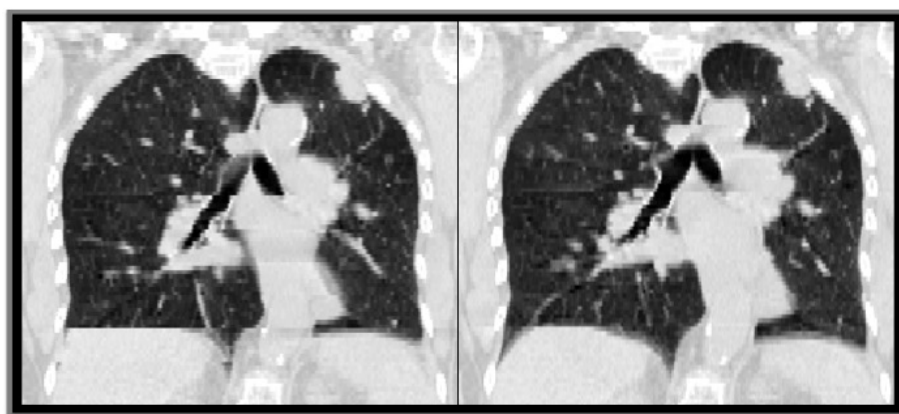


Figure 3.4. Sample Coronal Views for Sorting Comparison. Phase-sorted (left) and CM-sorted (right) clinical acquisition sample coronal views.

III. Statistical Analysis

i) Phase-Sorted Acquisitions

Significant differences were evident among the acquisition methods ($p < 0.024$). The mean CM values for the RM indicated a 30% increase in artifact presence relative to the clinical acquisition ($p < 0.001$). The data lacked evidence of a significant difference among clinical, GM, and OM acquisitions.

ii) CM-Sorted Acquisitions

Significant differences were evident among the acquisition methods test for CM sorting ($p < 0.0001$, F-test). The mean CM values for the RM indicated an increase in artifact presence relative to the clinical acquisition (37%; $p < 0.002$), the GM (26%; $p < 0.0052$), and OM (31%; $p < 0.001$). The data lacked evidence of a significant difference among clinical, GM, and OM acquisitions using CM sorting.

iii) CM vs Phase Sorting

CM sorting resulted in an estimated 24% reduction from phase sorting in mean CM (95% confidence interval=27-22). CM sorting applied to the OM resulted in a 27% reduction from the clinical standard of phase sorting applied to the clinical acquisition (95% confidence interval=34-20).

IV. Effective Dose Estimates

Figure 3.5 displays the estimated effective doses that were received by each patient for each acquisition method, with an adjustment for clinical tube current (100 mA) and clinical scan extent to the experimental acquisitions for proper dose comparison among all acquisition methods. The mean effective dose for the clinical method was $32.2 \pm 5.4 \text{ mSv}$, the gating method was $37.0 \pm 8.4 \text{ mSv}$, rescan method was $43.5 \pm 7.4 \text{ mSv}$, and oversampling method was $64.8 \pm 11.8 \text{ mSv}$. All experimental 4D CT acquisition methods, if applied clinically, would impart higher effective doses than the standard clinical 4D CT; the GM would impart the smallest dose increase and the OM would impart the largest dose increase.

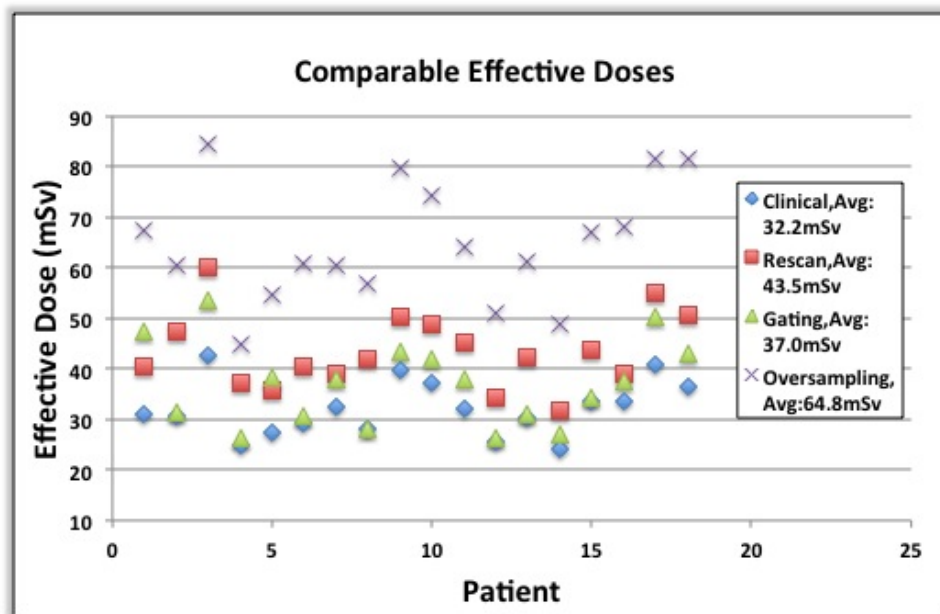


Figure 3.5. Estimated Cohort 1 Comparable Effective Doses. Effective doses with clinical tube current and clinical scan extent adjustment to experimental methods. The means of the dose estimates are shown in the legend.

V. Breathing Irregularities vs Artifact Presence

The relationship between the number of irregular breathing phases across the scan extent and the mean CM values are displayed for phase-sorted clinical and oversampling acquisitions (Figure 3.6), and for CM-sorted clinical and oversampling acquisitions (Figure 3.7) with the Pearson correlation coefficient shown in the upper right corners. Correlations between the phase-sorted (Figure 3.8) and CM-sorted (Figure 3.9) mean CM values versus percentage of breathing irregularities, and between the number of irregularities versus GM acquisition duration increase (Figure 3.10) are displayed below.

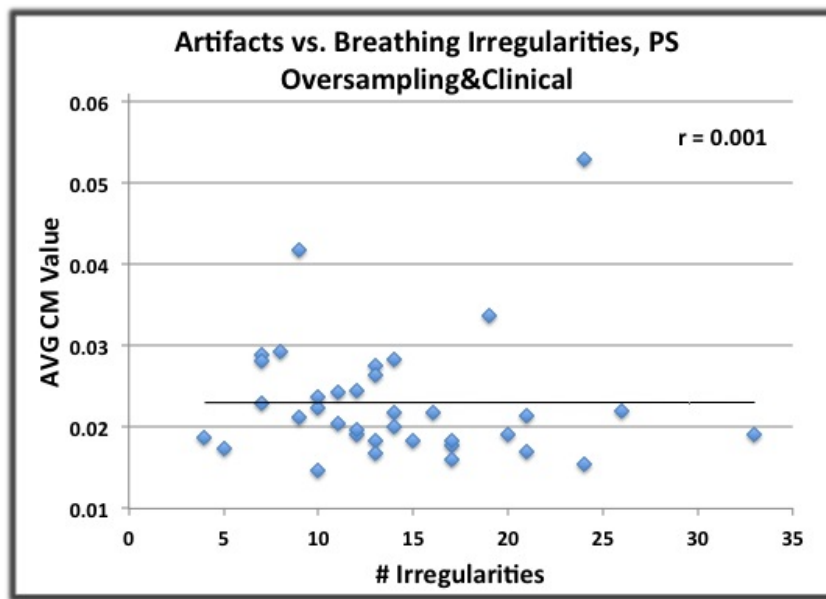


Figure 3.6. Phase-Sorted Artifacts vs Number of Breathing Irregularities.

Artifact presence in phase-sorted clinical and oversampling acquisitions, as defined by the mean CM values over component breathing phases, versus the number of breathing irregularities present during the scan extent. Pearson correlation coefficient is given by r in the upper right corner.

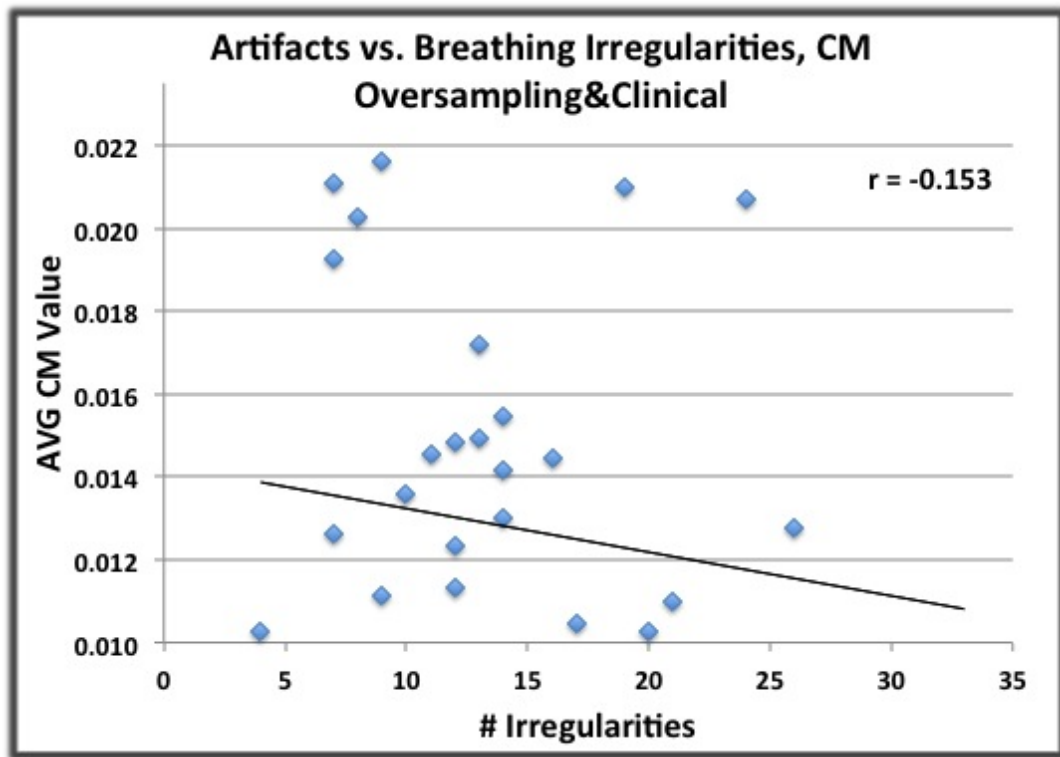


Figure 3.7. CM-Sorted Artifacts vs Number of Breathing Irregularities.

Artifact presence in CM-sorted clinical and oversampling acquisitions, as defined by the mean CM values over component breathing phases, versus the number of breathing irregularities present during the scan extent. Pearson correlation coefficient is given by r in the upper right corner.

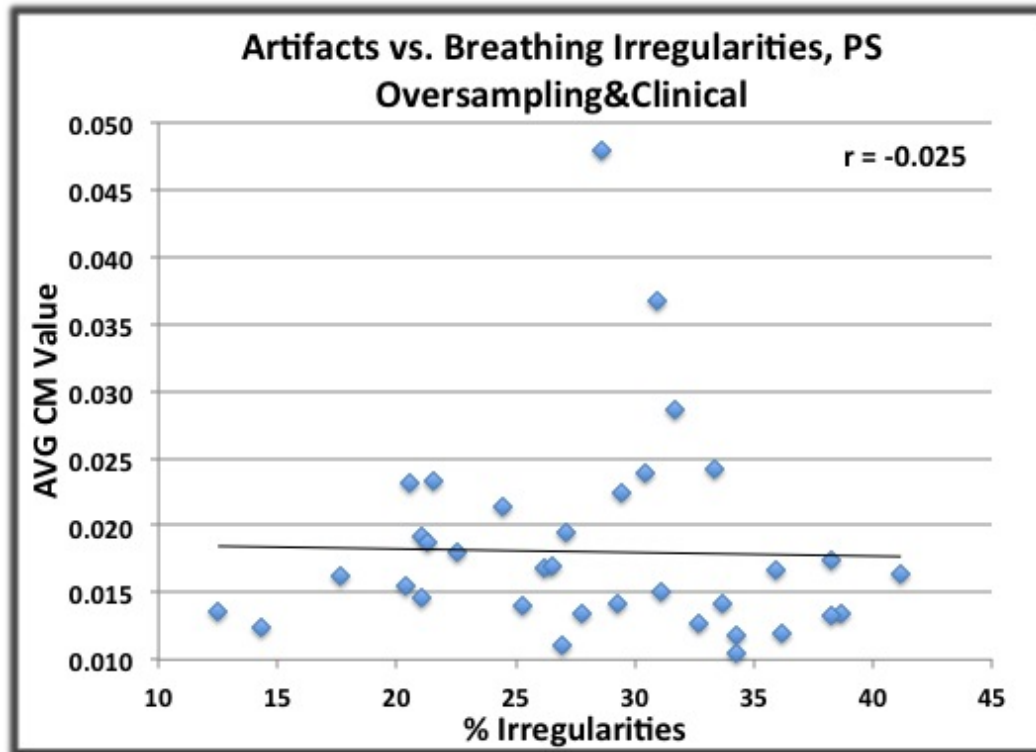


Figure 3.8. Phase-Sorted Artifacts vs Percentage of Breathing Irregularities.

Artifact presence in phase-sorted clinical and oversampling acquisitions, as defined by the mean CM values over component breathing phases, versus the percentage of breathing irregularities present during the scan extent. Pearson correlation coefficient is given by r in the upper right corner.

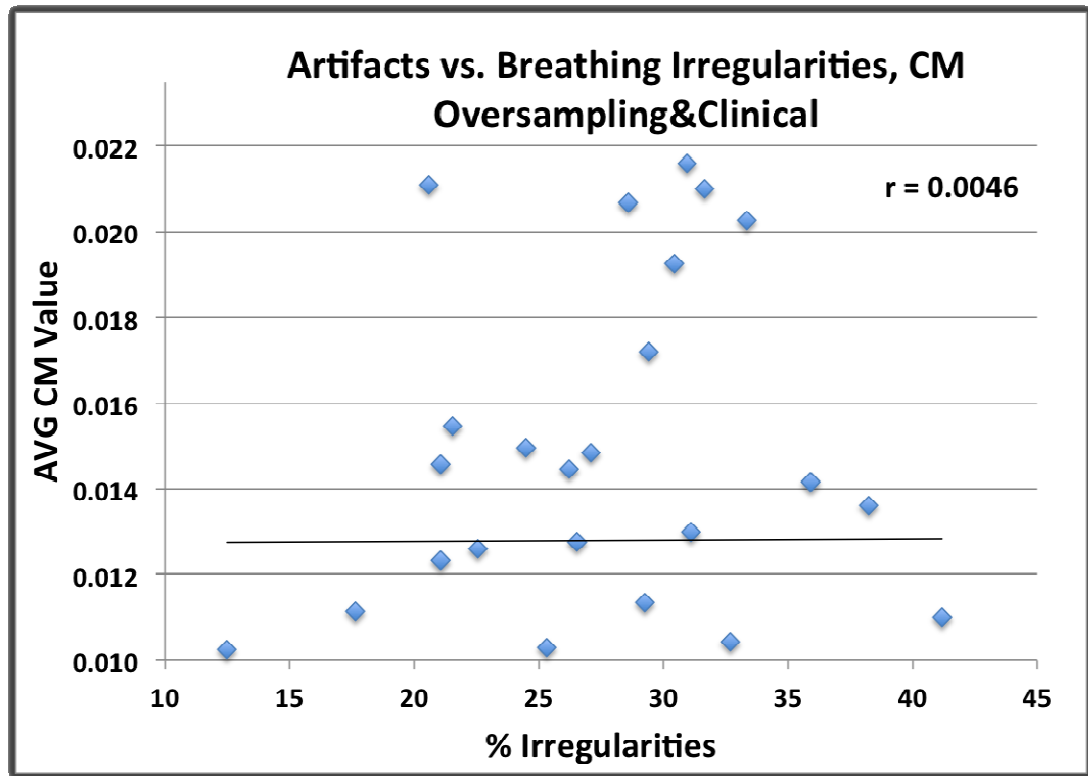


Figure 3.9. CM-Sorted Artifacts vs Percentage of Breathing Irregularities.

Artifact presence in CM-sorted clinical and oversampling acquisitions, as defined by the mean CM values over component breathing phases, versus the percentage of breathing irregularities present during the scan extent. Pearson correlation coefficient is given by r in the upper right corner.

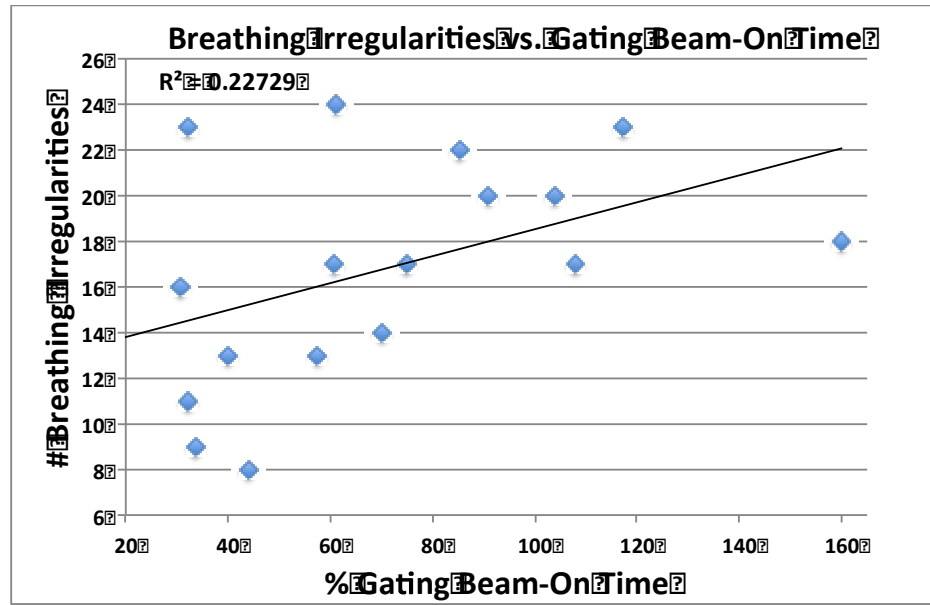


Figure 3.10. Gating Scan Duration vs Breathing Irregularities. Number of breathing irregularities versus percentage of GM beam-on time compared with the clinical acquisition beam-on time.

DISCUSSION

All CM-sorted acquisitions demonstrated a significant reduction in CM values from their phase-sorted equivalent, indicating that CM sorting produces superior image quality over phase sorting. Because we optimized an artifact-evaluating metric (CM) on possible image combinations per phase, we were able to minimize the presence of artifacts while retaining breathing information. This is consistent with previous findings from correlation-based image processing techniques.^(28, 59, 85, 86) This is the first study to demonstrate improvement in a prospective setting.

The CM-sorted OM achieved the greatest significant artifact reduction of 27% compared with the clinical acquisition with phase sorting. The OM contained the lowest artifact presence among CM-sorted acquisitions, but ranked poorly among phase-sorted

acquisitions. Phase sorting is limited because images are selected based on their distance in time to the phase definitions and do not incorporate image information for matching the anatomy, so the data increase in the OM may not have been fully used. Of all the experimental acquisition methods tested, the OM was the simplest to implement but also delivered the highest dose of all of the 4D CT acquisition methods tested.

The GM ranked well with phase sorting and moderately with CM sorting, but the artifact reduction was not significantly reduced relative to the clinical acquisition and the GM was roughly equal in image quality to the clinical acquisition for both sorting methods. More accurate detection of breathing irregularities and automatic x-ray beam control may improve this acquisition method. Because the respiratory trace could not be adequately viewed in its entirety in real time, only the last few cycles of the current breath determined the visual detection of an irregularity. A temporal lag also existed between the verbal indication to the operator to control the beam and the ensuing manual operation at the scanner by the operator. These factors led to a mean 54.6% of breathing irregularities that were not gated during image acquisition, which is a 26.7% reduction in acquired breathing irregularities relative to the clinical acquisition. Although more irregularities could have been gated, a larger artifact reduction was expected given the 27% reduction in acquired irregularities. A more robust irregularity quantification may be needed. It is also possible that the correlation between the breathing irregularities and the overall scan quality may not be as strong as previously thought, as indicated by the lack of correlation between mean CM values and irregularities (Results, section V). This lack of correlation indicated in there could be from a poor RPM correlation with internal anatomy, lack of a more robust irregularity detection method, or lack of an overall strong

correlation due to the averaging effects of higher and lesser magnitude artifacts within the 4D CT.

Stopping the beam for irregularities and waiting until regular breathing resumed increased the total scan duration. This was exacerbated by the couch position repetition in this implementation; leading to a mean 66.7% longer scan duration compared with the clinical acquisition. The scan duration seemed to be moderately correlated with the number of irregularities across the scan duration; this was expected, because more time is needed to acquire regular breathing for a patient with more irregular breaths. The GM acquisition duration would likely decrease with automatic control over the x-ray beam and evaluation of the real-time respiratory trace to determine when a full cycle of data had been acquired. However, even with this increase, all experimental and clinical 4D CT scans that were necessary for treatment simulation were obtained within the standard 60-minute clinical time slot.

The RM contained the greatest artifact presence of all the acquisitions and processing methods, and was significantly poorer than the clinical acquisition for both processing methods. The RM suffered from combining two independent data sets, and from not gating irregularities in the re-acquired images. The mean percentage of irregular breathing phases that occurred during the RM image re-acquisition was 28.4%. To evaluate the anatomical shifts present in the RM, the T0% phase of the RM was fused with the T0% phase of the clinical 4D CT for each patient. One of the 18 patients experienced a significant bone shift between the clinical 4D CT and RM, but most anatomical shifts were internal. Image segments were repeated without bias to scan

location, resulting in a mean of 37.8% of the re-acquired extent occurring in lung, with as few as a single couch position repetition in the lung.

The OM delivered at least twice the dose of that delivered by the clinical 4D CT method because of the increase in the cine duration needed to capture two cycles of breathing data. The RM delivered the second highest dose because a full clinical 4D CT was required in addition to the re-acquired images. The dose delivered with the RM could be reduced if the re-acquired couch positions were limited to those covering the lung. The GM delivered a dose equal to at least one clinical 4D CT scan because of the couch repetition inherent in the manual gating process, but the GM delivered the lowest overall effective dose of all the experimental methods tested. The effective dose delivered by the GM could be reduced by the exclusion of the couch repetition.

Because phase sorting selects image segments based on time, the images chosen could contain poor-quality data leaving the extra data available in the OM either not or only partially utilized. The CM sorting selects images based on time to phase in conjunction with spatial distributions present in the images, which opens the data selection to a potentially more accurate representation of breathing states. Because CM-sorted oversampling achieved the greatest reduction in artifact presence relative to the clinical standard, and because the OM has a simple and reproducible implementation, we consider it to be the best acquisition and sorting combination.

Although our study includes retrospective analysis of the two sorting methods, the focus was on alternative acquisition techniques, as these are lacking in the literature.^(32, 35-37, 44) We performed these acquisitions on a relatively large number of patients, without the use of simulations or phantoms.^(31, 39, 67) Artifacts were evaluated based on

quantitative assessments of the images, relying on a validation of the metric^(56, 87) and a general agreement with visual observation. The metric does not correlate exactly to visual assessment, but it provided a consistent method for evaluating scans in a relative fashion.

Another implementation weakness in this study is that the patients had to remain on the CT table in treatment position for at least 20 minutes longer than is needed for routine clinical scans while the experimental scans were acquired. To reduce anatomic shifts, the RM was acquired immediately after the clinical scan followed by the OM or GM, which were alternated from one patient to the next. Patients often became tired or uncomfortable with the longer scan times and the most promising scans were performed during the patient's worst state. Therefore we expect the scan quality to improve further if these methods are performed closer to when the patient first lies down on the table.

CONCLUSIONS

Artifact presence in the clinical 4D CT acquisition was compared with that in three experimental acquisition methods: data oversampling, beam-gating the breathing irregularities, and rescanning the clinical scan areas that were acquired during irregular breathing. Each was post-processed by the clinical standard phase sorting and by an alternative sorting method (CM sorting) that optimizes a correlation-based artifact metric. This alternative sorting method was found to significantly reduce artifact presence by approximately 24% across all acquisition methods. The oversampling acquisition combined with the alternative sorting method produced the highest statistically significant artifact reduction relative to the standard of phase-sorted clinical acquisition and was the simplest and most reproducible to implement.

CHAPTER 4

BEST 4D CT METHOD REPRODUCIBILITY

INTRODUCTION

The MD Anderson-approved protocol 2011-0631 included two patient cohorts; the first cohort of 18 patients was tested as described in Chapter 3 to determine the best 4D CT method, i.e., that which contained the lowest overall artifact presence. The objective for the second cohort was to determine the reproducibility of the identified best 4D CT method, and determine whether this best-chosen method consistently performs well across a broad patient sample when repeated. This Chapter describes our methods to test the reproducibility of the best-chosen 4D CT method, the CM-sorted oversampling acquisition (OCM).

I. The Patient Sample

The second cohort was comprised of 10 patients scheduled to undergo radiation therapy; half of the patients were women and half were men, and all had a clinical diagnosis of primary non-small cell lung cancer. The mean age of the participants was 68 ± 13.7 years. Each patient received a clinical 4D CT scan immediately followed by the chosen improved 4D CT method, which was repeated twice.

II. Improved Experimental Method

Each patient received a clinical phase-sorted 4D CT scan immediately followed by the three acquisitions of the improved experimental 4D CT method, termed

OCM1, OCM2, and OCM3 for each oversampling scan performed in chronologic order with CM sorting applied. After these four 4D CT scans, a normal-inhalation breath hold scan was acquired to serve as an anatomic reference. Not all patients could perform a stable normal inhalation breath hold and in those cases a deep inhalation breath hold was acquired. The OCM was performed as outlined in Chapter 3 (section III. iii), with a tube current of 50 mA in an effort to reduce the dose to the patient while maintaining acceptable image quality for the study. The scan extent was reduced from the clinical scan extent when possible to further reduce dose received by the patient. All other 4D CT acquisition parameters are the same as those outlined in Chapter 3 (section II).

III. Analysis

i) Artifact Presence

Statistical analysis consisted of a one-way mixed ANOVA to account for intra-patient correlation among the CM values across multiple phases. An F-test was used to test for an association with acquisition method and repetition. The resulting two-sided 95% confidence intervals are provided for each replicate scan for the percentage reduction in mean CM values for the OCM versus the clinical method. Inter-acquisition variability in CM values (inter-replicate deviation) was assessed by using the three independent acquisitions of oversampling in this sample of 10 patients. The resulting 95% limits of agreement were estimated using one-way mixed effects ANOVA.

ii) Image Quality vs. Patient Parameters

Given the wide range of percentage artifact improvement across patients, we explored possible relationships between image quality, patient characteristics, and breathing parameters. The CM values for the phase-sorted clinical method and OCM as well as the percentage improvement in OCM from the clinical method were used as the image quality parameters. Spearman correlation coefficients were calculated between all parameters. The Wilcoxon rank sum was used to test for significant differences between median CM values and the sex of the patient. Linear mixed regression was used to model the association between artifact presence and other variables. All tests were two-sided and p-values of 0.05 or less were considered statistically significant. Statistical analysis was carried out using SAS version 9.3 (SAS Institute, Cary, NC).

IV. Effective Dose Estimates

Effective dose estimates were calculated as outlined in Chapter 3 section VI. The estimated doses reported include an adjusted tube current (100 mA instead of 50 mA) and scan extent in the calculation to match the clinical values for appropriate comparison of experimental acquisitions to the clinical acquisition.

RESULTS

I. Patient Respiratory Statistics

Table 4.1 displays the summary statistics derived from the respiratory traces of the second cohort of patients. These statistics include percentage breathing

irregularities, mean abdominal displacement, mean breathing period, and mean number of breathing cycles per beam-on indicating the amount of data acquired at each couch position.

Acquisition Type	% Breathing Irregularities	Mean Displacement (cm)	Mean Breathing Period (s)	Mean Breathing Cycles/ Beam-On
Clinical	31.3 \pm 7.15	0.897 \pm 0.361	4.78 \pm 1.54	1.49 \pm 0.304
OCM1	25.4 \pm 6.79	0.771 \pm 0.302	4.43 \pm 1.17	3.03 \pm 0.599
OCM2	26.8 \pm 5.80	0.834 \pm 0.396	4.72 \pm 1.72	2.86 \pm 0.645
OCM3	30.0 \pm 5.84	0.783 \pm 0.299	4.60 \pm 1.39	2.93 \pm 0.646

Table 4.1. Cohort 2 Patient Respiratory Statistics. Percentage of breathing

irregularities, mean displacement between peak exhalation and peak inhalation, mean breathing period, and mean breathing cycles per beam-on. All parameters were calculated over the scan extent.

II. Analysis of Artifact Presence

Significant differences were evident among the clinical and OCM methods ($p < 0.0001$, F-test). The OCM resulted in an estimated 28% reduction in CM values from the clinical method (95% confidence interval = 20% - 37%; Figure 4.1). The inter-replicate deviation for the OCM 4D CT scans was within approximately $\pm 13\%$ of the cross acquisition average at the 0.05 significance level (Figure 4.2).^{(88,}

⁸⁹⁾ Figure 4.3 demonstrates the spread of CM values per breathing phase over each acquisition method per cohort 2 patients. Examples of coronal views of the clinical phase-sorted 4D CT and the three subsequent oversampling CM-sorted 4D CT scans are given in Figure 4.4.

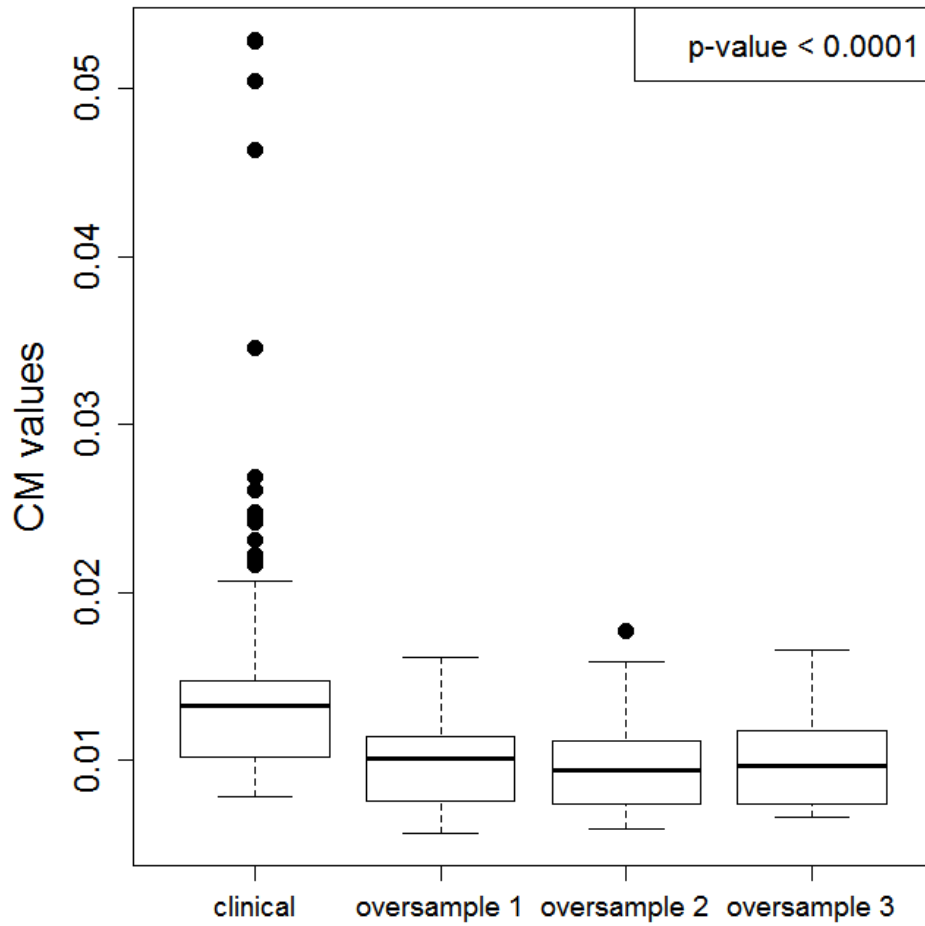


Figure 4.1. Boxplot of CM Values by Acquisition Method. The sum of the absolute value of the CM values over all couch positions per breathing phase per acquisition method.

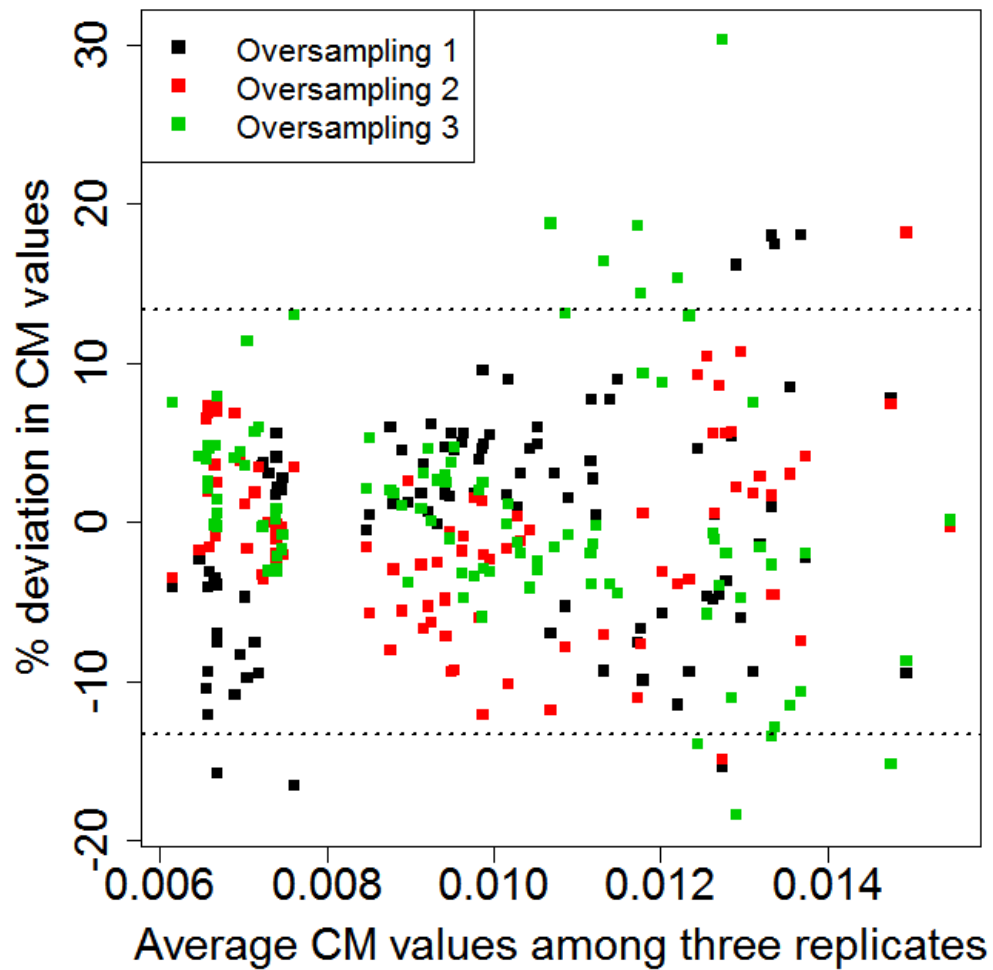


Figure 4.2. Bland-Altman Plot for Oversampling CM Value Inter-Acquisition

Agreement. Observed and expected percentage deviation from the mean CM values under the three oversampling acquisitions. One-way mixed effects ANOVA obtains 95% confidence boundaries = $\pm 13\%$.

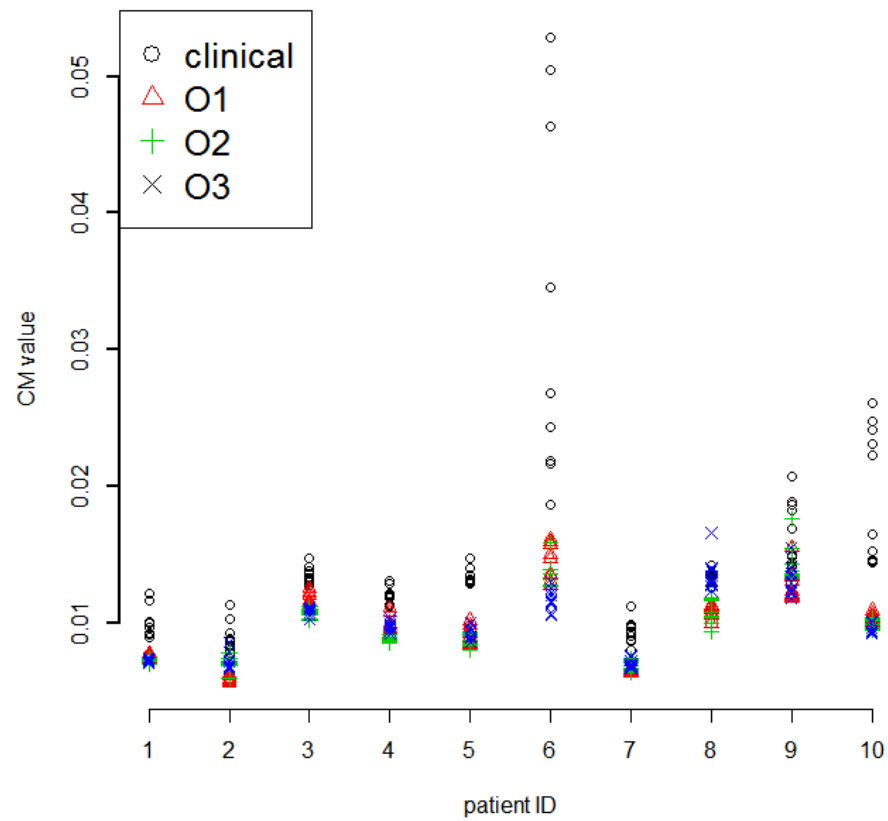


Figure 4.3. CM Value Spread per Patient. Spread of CM values per breathing phase over all acquisitions per patient.

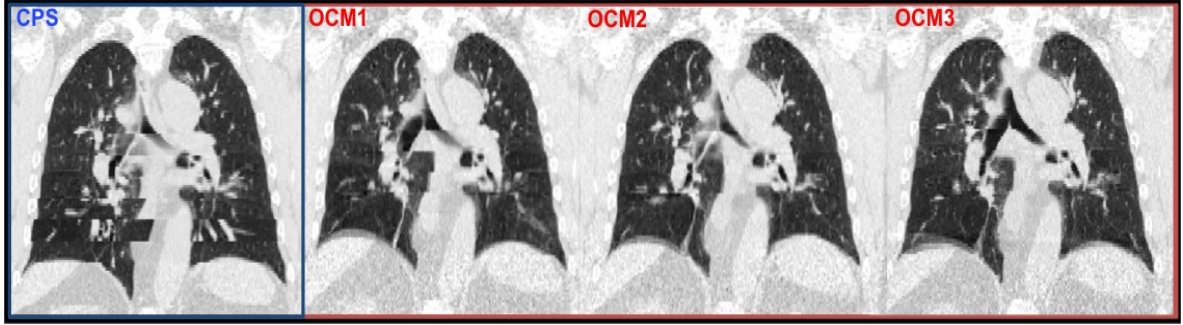


Figure 4.4. Coronal views of the clinical and repeat OCM scans. Sample images from the clinical phase-sorted method (CPS), and three CM-sorted oversampling scans (OCM1, OCM2, OCM3) from a patient case.

III. Image Quality vs. Patient Parameters

Spearman correlation coefficients between CM values and breathing parameters are given in Table 4.2. Breathing parameters include means over the scan extent: breathing cycles per couch position (amount of data collected), breathing period, abdominal displacement, percentage breathing irregularities, and number of breathing irregularities. Spearman correlation between percentage improvement in CM values for OCM from phase-sorted clinical CM values and breathing parameters are given in Table 4.3.

A Wilcoxon rank sum test was performed to evaluate the significance of association of image quality with patient sex. Percentage improvement from clinical to the OCM did not associate significantly with sex, however CM values for both methods were significantly associated (clinical median CM values (F/M) = 0.013/0.017; p -value<0.007, OCM median CM values (F/M) = 0.009/0.012; p -value<0.005) with men demonstrating higher artifact presence generally. Results of the mixed regression analysis for association with percentage improvement in

CM values indicated a significant association with age (p-value< 0.035), indicating that older patients experienced greater improvement from the OCM. Significant correlations between improvement in image quality and OCM, breathing period, age, and amount of data are demonstrated in Figure 4.5.

4D CT Method		Cycles/Couch	PER	DIS	%IRREG	IRREG
Clinical CM Values (N=28)	ρ	-0.365	0.407	0.239	-0.108	-0.213
	p-value	0.057	0.032	0.221	0.585	0.276
OCM CM Values (N=48)	ρ	-0.335	0.296	0.0740	-0.148	-0.270
	p-value	0.020	0.041	0.616	0.315	0.063

Table 4.2. Spearman correlation coefficients between CM values and breathing parameters.

Cycles/Couch = breathing cycles per couch position; PER = breathing period; DIS = abdominal displacement; %IRREG = percentage breathing irregularities; IRREG = number of breathing irregularities

% Difference from clinical <input type="checkbox"/> <input type="checkbox"/> oversampling					
% CM Improvement	Cycles/Couch	PER	DIS	%IRREG	IRREG
ρ	0.490	0.375	0.177	-0.150	-0.215
p-value	0.0004	0.009	0.228	0.308	0.142

Table 4.3. Spearman correlation coefficient between percentage improvement in CM values with OCM and percentage difference in breathing parameters.

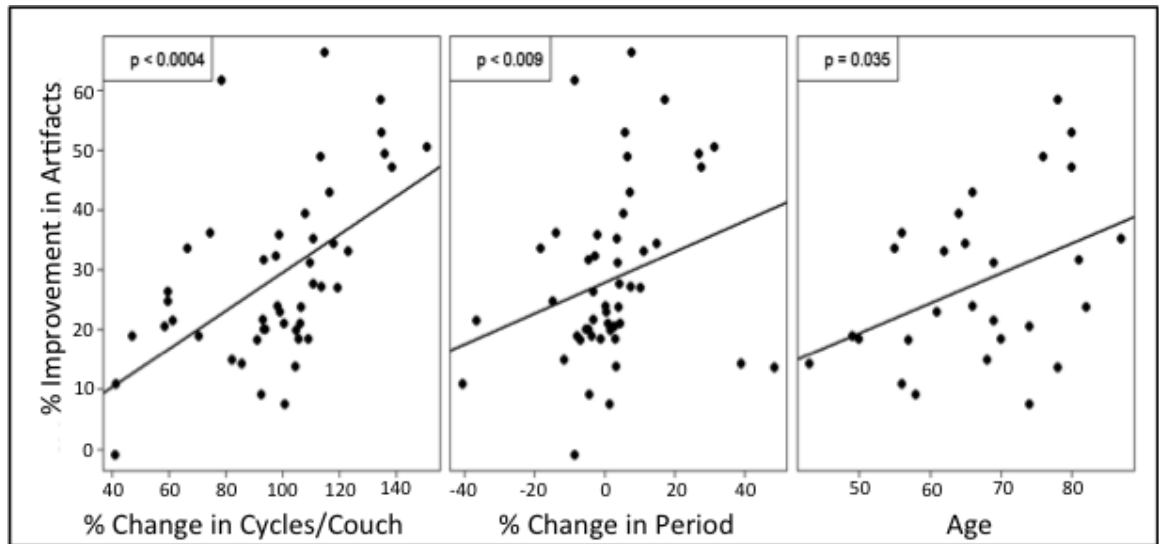


Figure 4.5. Significant patient parameters indicative of image quality.

Percentage improvement with OCM versus percentage difference in breathing cycles per couch position (left), percentage difference in breathing period (middle), and age (right).

IV. Effective Dose Estimates

Figure 4.6 displays the estimated effective doses that were received by each patient in the second cohort for each 4D CT acquisition, with an adjustment for clinical tube current (100 mA) and clinical scan extent to the OM scans for proper dose comparison with the clinical dose received. The mean of the effective dose for the clinical method was 31.5 ± 4.3 mSv, for the first OM scan, it was 60.4 ± 8.6 mSv, for the second OM scan, 59.4 ± 9.8 mSv, and for the third OM scan, 60.1 ± 9.5 mSv. All OM acquisitions, if applied clinically, would impart approximately twice the effective dose of the standard clinical 4D CT.

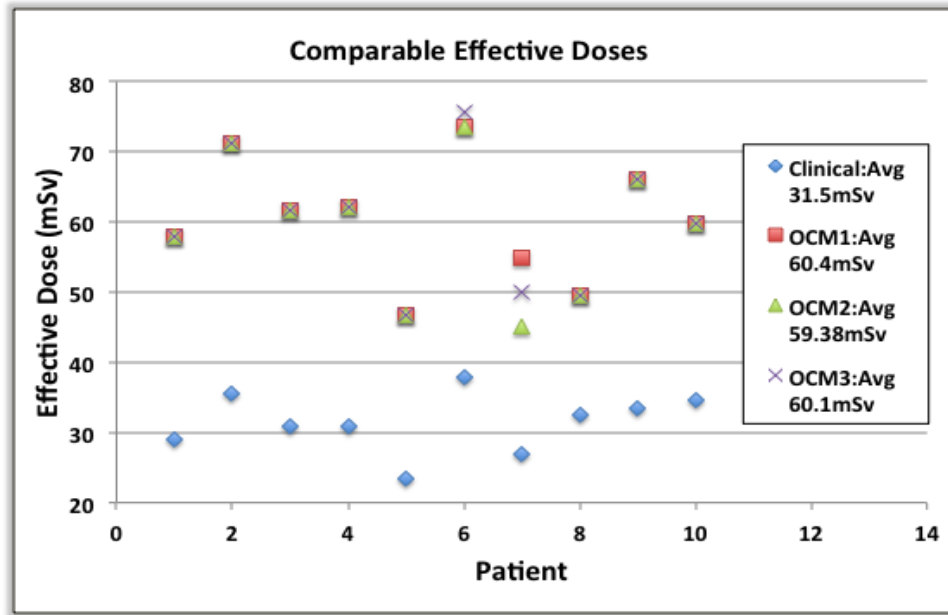


Figure 4.6 Estimated Comparable Effective Doses for Cohort 2. Estimated effective doses (mSv) received per 4D CT acquisition for the 10 patients in cohort 2. Effective doses for oversampling scans (O1, O2, O3) include clinical tube current and scan extent settings.

DISCUSSION

A significant reduction in artifacts relative to the clinical phase-sorted 4D CT was again reached with the CM-sorted oversampling method for the second patient cohort. The estimated artifact reduction based on CM values was 28%, a 1% improvement from the 27% reduction achieved with the first patient cohort. This finding indicates that the CM-sorted oversampling 4D CT method was consistent among a larger patient sample. The inter-replicate deviation for the repeated OCM scans was 13%; this is fairly tight, especially considering the outlier cases, patients six and 10. These outliers result in part from a patient effect, in which certain patients generally have higher CM values for any

acquisition specific to them, and they are also reflect irregular breathing that caused high magnitude artifacts that dramatically decreased the overall quality of the scan. These patients generally experienced a larger percentage improvement in artifact presence with the OCM because the image-guided sorting and extra available data allowed avoidance of the image segments that were acquired during irregular breathing. The OCM estimated effective doses are consistent, varying slightly because of changes in cine duration, and are roughly double the dose of the clinical acquisition.

Several statistically significant associations between artifact presence (CM values) and patient parameters were found, although the strength of the correlation was moderate at best. Higher numbers of breathing cycles per couch position yielded lower CM values (fewer artifacts) for the OCM, whereas shorter breathing periods yielded fewer artifacts for both clinical and OCM methods, and women tended to have fewer artifacts than men for both methods. We are unsure why women's images would exhibit less artifacts than men's images, a possible explanation is a difference in chest and abdominal breathing distribution between sexes, or perhaps multiple factors pertaining to the patient's health are present. A higher percentage improvement with the OCM 4D CT was significantly correlated with patient age, breathing period, and breathing cycles acquired per couch position.

In general, the older patients experienced more improvement with the OCM 4D CT than younger patients, although no significant correlation was noted between age and CM values. This may reflect an increased discomfort experienced by the older patients during the scan session, resulting in more variable breathing; the image-guided CM-sorting and additional available data in the oversampling allow more accurate images to be chosen in

these situations. This would not be reflected well in our irregularity quantification, which did not significantly correlate with image quality, although various reports cite the negative effect of breathing irregularities on image quality (as cited in the introduction to Chapter 2. The irregularity quantification calculated does not characterize all of the points in the respiratory trace, just 10 points of the 150 points present during a 5 s breathing period. Another issue with the breathing quantification is the lack of a reference trace or relative parameter indicative of quality across multiple scans per patient. Each scan's irregularities are identified as irregular with regard to that particular scan's respiratory trace; this makes it difficult to compare irregularities present across multiple scans per patient.

Patients with faster breathing (i.e., a shorter breathing period) tended to have less artifact presence in their clinical and OCM 4D CT scans, as well as a higher percentage improvement with the OCM scans. This may be from smaller motion experienced with shorter breaths, or from an increase in breathing cycles per couch position collected because of the faster breathing. There was a significant correlation between breathing period and breathing cycles per couch position for the clinical and OCM methods (-0.91, p -value<0.0001 clinical, -0.618, p -value<0.0005 OCM) indicating that the amount of data that are collected may be driving the correlation with the breathing period. An increase in breathing cycles per couch position reflects the amount of data acquired, and the more data collected, the fewer the artifacts with image-guided sorting. Phase sorting is limited to selecting image segments solely based on time, and therefore does not show an improvement in image quality with added data; the data may be there, but are not being used.

Patient sex was significantly correlated with artifact presence in both 4D CT methods, with 4D CT scans for women having lower CM values than for men. This may reflect an inherent difference in breathing between men and women that was reflected in our evaluation of image quality. Differences in breathing through the nose or mouth have been reported to affect breathing variability for both men and women, albeit in different manners.⁽⁹⁰⁾ It has been reported that men tend to experience a larger AP abdominal displacement during quiet breathing⁽⁹¹⁾ as well as larger abdominal displacements during deep breathing.⁽⁹²⁾ There are many factors that affect breathing characteristics for both sex and age, and these correlations were moderate at best, so these non-zero correlations may also reflect other anatomic parameters or medical issues not explored in this work.

Yamamoto et al.⁽²⁶⁾ evaluated correlations between similar parameters and visually identified artifacts, and found significant correlations between image quality, abdominal displacement, and breathing period. We also experienced a significant correlation with breathing period, but the correlation values in the Yamamoto study were not reported, and a moderate correlation may also have been present. Their correlation was not based on an overall measure of image quality however, but on a fraction of cases evaluated that contained any artifact at all. We did not calculate a significant correlation between abdominal displacement and artifact presence for either 4D CT method. Because the image sampling and number of available image segments depends on time and the motion happening within that time, perhaps the artifact presence is related more to the period than the displacement.

CONCLUSION

A second cohort of 10 patients with primary thoracic cancer was used to test the reproducibility of the artifact reduction achieved with the CM-sorted oversampling 4D CT method. The oversampling method achieved a 28% artifact reduction compared with the clinical method, a 1% difference from the first cohort, and demonstrated a low inter-replicate deviation. Therefore the CM-sorted oversampling 4D CT method consistently and significantly reduced the artifact presence relative to the clinical phase-sorted standard. In general, larger reductions in artifacts with the improved method were obtained for patients who breathed more quickly, were more elderly, and had scans with greater amounts of data collected.

References

1. Ehrhardt J, Werner R, Saring D, Frenzel T, Lu W, Low D, Handels H. An optical flow based method for improved reconstruction of 4D CT data sets acquired during free breathing. *Med Phys.* 2007;34(2):711-21. Epub 2007/03/29. PubMed PMID: 17388189.
2. Ritchie CJ, Godwin JD, Crawford CR, Stanford W, Anno H, Kim Y. Minimum scan speeds for suppression of motion artifacts in CT. *Radiology.* 1992;185(1):37-42. Epub 1992/10/01. PubMed PMID: 1523332.
3. Tarver RD, Conces DJ, Jr., Godwin JD. Motion artifacts on CT simulate bronchiectasis. *AJR Am J Roentgenol.* 1988;151(6):1117-9. Epub 1988/12/01. PubMed PMID: 3263763.
4. Rietzel E, Pan T, Chen GT. Four-dimensional computed tomography: image formation and clinical protocol. *Med Phys.* 2005;32(4):874-89. Epub 2005/05/18. PubMed PMID: 15895570.
5. Low DA, Nystrom M, Kalinin E, Parikh P, Dempsey JF, Bradley JD, Mutic S, Wahab SH, Islam T, Christensen G, Politte DG, Whiting BR. A method for the reconstruction of four-dimensional synchronized CT scans acquired during free breathing. *Med Phys.* 2003;30(6):1254-63. Epub 2003/07/11. PubMed PMID: 12852551.
6. Keall PJ, Mageras GS, Balter JM, Emery RS, Forster KM, Jiang SB, Kapatoes JM, Kubo HD, Low DA, Murphy MJ, Murray BR, Ramsey CR, Herk MBv, Vedam SS, Wong JW, Yorke E. The Management of Respiratory Motion in Radiation Oncology. Report 91. 2006.
7. Guckenberger M, Wilbert J, Meyer J, Baier K, Richter A, Flentje M. Is a Single Respiratory Correlated 4D-CT Study Sufficient for Evaluation of Breathing Motion? . *International Journal of Radiation Oncology Biology Physics.* 2007;67(5):1352-9.
8. Vedam SS, Keall PJ, Kini VR, Mostafavi H, Shukla HP, Mohan R. Acquiring a four-dimensional computed tomography dataset using an external respiratory signal. *Phys Med Biol.* 2003;48(1):45-62. Epub 2003/02/05. PubMed PMID: 12564500.

9. Pan T, Lee TY, Rietzel E, Chen GT. 4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT. *Med Phys*. 2004;31(2):333-40. Epub 2004/03/06. PubMed PMID: 15000619.
10. Keall PJ, Joshi S, Vedam SS, Siebers JV, Kini VR, Mohan R. Four-dimensional radiotherapy planning for DMLC-based respiratory motion tracking. *Med Phys*. 2005;32(4):942-51. Epub 2005/05/18. PubMed PMID: 15895577.
11. Balter JM, Ten Haken RK, Lawrence TS, Lam KL, Robertson JM. Uncertainties in CT-based radiation therapy treatment planning associated with patient breathing. *Int J Radiat Oncol Biol Phys*. 1996;36(1):167-74. Epub 1996/08/01. PubMed PMID: 8823272.
12. Ezhil M, Vedam S, Balter P, Choi B, Mirkovic D, Starkschall G, Chang JY. Determination of patient-specific internal gross tumor volumes for lung cancer using four-dimensional computed tomography. *Radiat Oncol*. 2009;4:4. doi: 10.1186/1748-717X-4-4. PubMed PMID: 19173738; PubMed Central PMCID: PMC2645420.
13. Glide-Hurst CK, Schwenker Smith M, Ajlouni M, Chetty IJ. Evaluation of two synchronized external surrogates for 4D CT sorting. *J Appl Clin Med Phys*. 2013;14(6):4301. doi: 10.1120/jacmp.v14i6.4301. PubMed PMID: 24257273.
14. Li XA, Stepaniak C, Gore E. Technical and dosimetric aspects of respiratory gating using a pressure-sensor motion monitoring system. *Med Phys*. 2006;33(1):145-54. PubMed PMID: 16485421.
15. M Wilms RW, J Ehrhardt, A Schmidt-Richberg, H-P Schlemmer, and H Handels. Multivariate regression approaches for surrogate-based diffeomorphic estimation of respiratory motion in radiation therapy. *Phys Med Biol*. 2014;59(5):1147 - 64

16. Zamora DA, Riegel AC, Sun X, Balter P, Starkschall G, Mawlawi O, Pan T. Thoracic target volume delineation using various maximum-intensity projection computed tomography image sets for radiotherapy treatment planning. *Med Phys.* 2010;37(11):5811-20. PubMed PMID: 21158293; PubMed Central PMCID: PMC3810265.
17. Wu H, Zhao Q, Berbeco RI, Nishioka S, Shirato H, Jiang SB. Gating based on internal/external signals with dynamic correlation updates. *Phys Med Biol.* 2008;53(24):7137-50. doi: 10.1088/0031-9155/53/24/009. PubMed PMID: 19033643.
18. Vedam SS, Kini VR, Keall PJ, Ramakrishnan V, Mostafavi H, Mohan R. Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker. *Med Phys.* 2003;30(4):505-13. Epub 2003/05/02. PubMed PMID: 12722802.
19. Vedam SS, Keall PJ, Kini VR, Mohan R. Determining parameters for respiration-gated radiotherapy. *Med Phys.* 2001;28(10):2139-46. Epub 2001/11/07. PubMed PMID: 11695776.
20. Jerrold T. Bushberg JAS, Edwin M. Leidholdt, John M. Boone. *The Essential Physics of Medical Imaging*. 3rd ed: Lippincott Williams and Wilkins; 2011. 1048 p.
21. Pan T. Comparison of helical and cine acquisitions for 4D-CT imaging with multislice CT. *Med Phys.* 2005;32(2):627-34. Epub 2005/03/26. PubMed PMID: 15789609.
22. Ford E, Mageras G, York E, Ling C. Respiration-correlated spiral CT: A method of measuring respiratory-induced anatomic motion for radiation treatment planning. *Med Phys.* 2002;29(6):1340-. PubMed PMID: ISI:000176373400694.
23. Keall PJ, Starkschall G, Shukla H, Forster KM, Ortiz V, Stevens CW, Vedam SS, George R, Guerrero T, Mohan R. Acquiring 4D thoracic CT scans using a multislice helical method. *Phys Med Biol.* 2004;49(10):2053-67. Epub 2004/06/25. PubMed PMID: 15214541.
24. Joseph V. Hajnal DLGH, David J. Hawkes. *Medical Image Registration*. Neuman MR, editor: CRC Press; 2001.

25. Barrett JF, Keat N. Artifacts in CT: recognition and avoidance. *Radiographics*. 2004;24(6):1679-91. Epub 2004/11/13. doi: 10.1148/rg.246045065. PubMed PMID: 15537976.
26. Yamamoto T, Langner U, Loo BW, Jr., Shen J, Keall PJ. Retrospective analysis of artifacts in four-dimensional CT images of 50 abdominal and thoracic radiotherapy patients. *Int J Radiat Oncol Biol Phys*. 2008;72(4):1250-8. Epub 2008/10/01. doi: 10.1016/j.ijrobp.2008.06.1937. PubMed PMID: 18823717; PubMed Central PMCID: PMC2583232.
27. Watkins WT, Li R, Lewis J, Park JC, Sandhu A, Jiang SB, Song WY. Patient-specific motion artifacts in 4DCT. *Med Phys*. 2010;37(6):2855-61. Epub 2010/07/17. PubMed PMID: 20632597.
28. Carnes G, Gaede S, Yu E, Van Dyk J, Battista J, Lee TY. A fully automated non-external marker 4D-CT sorting algorithm using a serial cine scanning protocol. *Phys Med Biol*. 2009;54(7):2049-66. doi: 10.1088/0031-9155/54/7/013. PubMed PMID: 19287079.
29. Jie Wei GL. Quantification of motion artifacts in 4DCT using global Fourier analysis. *Signal Processing in Medicine and Biology Symposium (SPMB)*, 2012 IEEE; December 1, 2012. p. 1-5.
30. Abdelnour AF, Nehmeh SA, Pan T, Humm JL, Vernon P, Schoder H, Rosenzweig KE, Mageras GS, Yorke E, Larson SM, Erdi YE. Phase and amplitude binning for 4D-CT imaging. *Phys Med Biol*. 2007;52(12):3515-29. Epub 2007/08/01. doi: 10.1088/0031-9155/52/12/012. PubMed PMID: 17664557.
31. Keall PJ, Vedam SS, George R, Williamson JF. Respiratory regularity gated 4D CT acquisition: concepts and proof of principle. *Australas Phys Eng Sci Med*. 2007;30(3):211-20. Epub 2007/11/30. PubMed PMID: 18044305.
32. Pan T, Sun X, Luo D. Improvement of the cine-CT based 4D-CT imaging. *Med Phys*. 2007;34(11):4499-503. Epub 2007/12/13. PubMed PMID: 18072515.
33. Fitzpatrick MJ, Starkschall, George, Antolak, John A., Fu, Jun, Shukla Himanshu, Keall, Paul J., Klahr, Paul, Mohan, Radhe. Displacement-based binning of time-dependent computed tomography image data sets. *Med Phys*. 2006;33(1).

34. Rietzel E, Chen GT. Improving retrospective sorting of 4D computed tomography data. *Med Phys*. 2006;33(2):377-9. Epub 2006/03/15. PubMed PMID: 16532943.
35. Zhang Y, Yang J, Zhang L, Court LE, Balter PA, Dong L. Modeling respiratory motion for reducing motion artifacts in 4D CT images. *Med Phys*. 2013;40(4):041716. doi: 10.1118/1.4795133. PubMed PMID: 23556886.
36. Gianoli C, Riboldi M, Spadea MF, Travaini LL, Ferrari M, Mei R, Orecchia R, Baroni G. A multiple points method for 4D CT image sorting. *Med Phys*. 2011;38(2):656-67. PubMed PMID: 21452703.
37. Hertanto A, Zhang Q, Hu Y, Dzyubak O. Reduction of Irregular breathing artifacts in respiration-correlated CT images using a respiratory motion model. *Med Phys*. 2012;39(6):3070-8.
38. Lu W, Parikh PJ, Hubenschmidt JP, Bradley JD, Low DA. A comparison between amplitude sorting and phase-angle sorting using external respiratory measurement for 4D CT. *Med Phys*. 2006;33(8):2964-74. Epub 2006/09/13. PubMed PMID: 16964875.
39. Langner UW, Keall PJ. Accuracy in the localization of thoracic and abdominal tumors using respiratory displacement, velocity, and phase. *Med Phys*. 2009;36(2):386-93. Epub 2009/03/19. PubMed PMID: 19291977; PubMed Central PMCID: PMC2736730.
40. Schreiber E, Chen GT, Xing L. Image interpolation in 4D CT using a BSpline deformable registration model. *Int J Radiat Oncol Biol Phys*. 2006;64(5):1537-50. Epub 2006/03/01. doi: 10.1016/j.ijrobp.2005.11.018. PubMed PMID: 16503382.
41. Guckenberger M, Weininger M, Wilbert J, Richter A, Baier K, Krieger T, Polat B, Flentje M. Influence of retrospective sorting on image quality in respiratory correlated computed tomography. *Radiother Oncol*. 2007;85(2):223-31. Epub 2007/09/15. doi: 10.1016/j.radonc.2007.08.002. PubMed PMID: 17854931.

42. Han D, Bayouth J, Song Q, Bhatia S, Sonka M, Wu X. Feature Guided Motion Artifact Reduction with Structure-Awareness in 4D CT Images. *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2011;2011:1057-64. doi: 10.1109/CVPR.2011.5995561. PubMed PMID: 22058647; PubMed Central PMCID: PMC3207360.
43. Johnston RG, Yen EK. The Ineffectiveness of the Correlation Coefficient for Image Comparisons. *Journal of Physical Security*. 2007;2(1).
44. Olsen JR, Lu W, Hubenschmidt JP, Nystrom MM, Klahr P, Bradley JD, Low DA, Parikh PJ. Effect of novel amplitude/phase binning algorithm on commercial four-dimensional computed tomography quality. *Int J Radiat Oncol Biol Phys*. 2008;70(1):243-52. Epub 2007/11/27. doi: 10.1016/j.ijrobp.2007.09.013. PubMed PMID: 18037590; PubMed Central PMCID: PMC2702992.
45. Wei JL, G. Quantification of motion artifacts in 4DCT using Global Fourier Analysis. *Signal Processing in Medicine and Biology Symposium (SPMB)*, 2012 IEEE2012.
46. Langner UW, Keall PJ. Quantification of artifact reduction with real-time cine four-dimensional computed tomography acquisition methods. *Int J Radiat Oncol Biol Phys*. 2010;76(4):1242-50. Epub 2009/11/27. doi: 10.1016/j.ijrobp.2009.07.013. PubMed PMID: 19939579.
47. Langner UW, Keall PJ. Prospective displacement and velocity-based cine 4D CT. *Med Phys*. 2008;35(10):4501-12. Epub 2008/11/04. PubMed PMID: 18975697; PubMed Central PMCID: PMC2736757.
48. Coolens C, Bracken J, Driscoll B, Hope A, Jaffray D. Dynamic volume vs respiratory correlated 4DCT for motion assessment in radiation therapy simulation. *Med Phys*. 2012;39(5):2669-81. doi: 10.1118/1.4704498. PubMed PMID: 22559637.
49. Persson GF, Nygaard DE, Af Rosenschold PM, Richter Vogelius I, Josipovic M, Specht L, Korreman SS. Artifacts in conventional computed tomography (CT) and free breathing four-dimensional CT induce uncertainty in gross tumor volume determination. *Int J Radiat Oncol Biol Phys*. 2011;80(5):1573-80. Epub 2010/12/18. doi: 10.1016/j.ijrobp.2010.10.036. PubMed PMID: 21163584.

50. Persson GF, Nygaard DE, Brink C, Jahn JW, Munck af Rosenschold P, Specht L, Korreman SS. Deviations in delineated GTV caused by artefacts in 4DCT. *Radiother Oncol.* 2010;96(1):61-6. Epub 2010/06/24. doi: 10.1016/j.radonc.2010.04.019. PubMed PMID: 20570002.
51. Mori S, Endo M, Komatsu S, Yashiro T, Kandatsu S, Baba M. Four-dimensional measurement of lung tumor displacement using 256-multi-slice CT-scanner. *Lung cancer.* 2007;56(1):59-67. doi: 10.1016/j.lungcan.2006.11.011. PubMed PMID: 17157951.
52. Mori S, Hara R, Yanagi T, Sharp GC, Kumagai M, Asakura H, Kishimoto R, Yamada S, Kandatsu S, Kamada T. Four-dimensional measurement of intrafractional respiratory motion of pancreatic tumors using a 256 multi-slice CT scanner. *Radiother Oncol.* 2009;92(2):231-7. doi: 10.1016/j.radonc.2008.12.015. PubMed PMID: 19211167.
53. Wink N, Panknin C, Solberg TD. Phase versus amplitude sorting of 4D-CT data. *J Appl Clin Med Phys.* 2006;7(1):77-85. Epub 2006/03/07. PubMed PMID: 16518319.
54. Lu W, Parikh PJ, Hubenschmidt JP, Bradley JD, Low DA. A comparison between amplitude sorting and phase-angle sorting using external respiratory measurement for 4D CT. *Med Phys.* 2006;33(8):2964-74. PubMed PMID: 16964875.
55. Han D, Bayouth J, Bhatia S, Sonka M, Wu X. Characterization and identification of spatial artifacts during 4D-CT imaging. *Med Phys.* 2011;38(4):2074-87. Epub 2011/06/02. PubMed PMID: 21626940; PubMed Central PMCID: PMC3078159.
56. Cui G, Jew B, Hong JC, Johnston EW, Loo BW, Jr., Maxim PG. An automated method for comparing motion artifacts in cine four-dimensional computed tomography images. *J Appl Clin Med Phys.* 2012;13(6):3838. Epub 2012/11/15. doi: 10.1120/jacmp.v13i6.3838. PubMed PMID: 23149777.
57. Vasquez AC, Runz A, Echner G, Sroka-Perez G, Karger CP. Comparison of two respiration monitoring systems for 4D imaging with a Siemens CT using a new dynamic breathing phantom. *Phys Med Biol.* 2012;57(9):N131-43. doi: 10.1088/0031-9155/57/9/N131. PubMed PMID: 22504160.

58. Mutaf YD, Antolak JA, Brinkmann DH. The impact of temporal inaccuracies on 4DCT image quality. *Med Phys*. 2007;34(5):1615-22. Epub 2007/06/09. PubMed PMID: 17555243.
59. Johnston E, Diehn M, Murphy JD, Loo BW, Jr., Maxim PG. Reducing 4D CT artifacts using optimized sorting based on anatomic similarity. *Med Phys*. 2011;38(5):2424-9. PubMed PMID: 21776777.
60. Pierce G, Wang K, Battista J, Lee TY. Applying an animal model to quantify the uncertainties of an image-based 4D-CT algorithm. *Phys Med Biol*. 2012;57(11):3571-84. doi: 10.1088/0031-9155/57/11/3571. PubMed PMID: 22588201.
61. McClelland JR, Hughes S, Modat M, Qureshi A, Ahmad S, Landau DB, Ourselin S, Hawkes DJ. Inter-fraction variations in respiratory motion models. *Phys Med Biol*. 2011;56(1):251-72. doi: 10.1088/0031-9155/56/1/015. PubMed PMID: 21149951.
62. Wolthaus JW, Sonke JJ, van Herk M, Damen EM. Reconstruction of a time-averaged midposition CT scan for radiotherapy planning of lung cancer patients using deformable registration. *Med Phys*. 2008;35(9):3998-4011. PubMed PMID: 18841851.
63. Guerrero T, Sanders K, Noyola-Martinez J, Castillo E, Zhang Y, Tapia R, Guerra R, Borghero Y, Komaki R. Quantification of regional ventilation from treatment planning CT. *Int J Radiat Oncol Biol Phys*. 2005;62(3):630-4. doi: 10.1016/j.ijrobp.2005.03.023. PubMed PMID: 15936537.
64. Simon BA. Non-invasive imaging of regional lung function using x-ray computed tomography. *Journal of clinical monitoring and computing*. 2000;16(5-6):433-42. PubMed PMID: 12580227.
65. McClelland JR, Blackall JM, Tarte S, Chandler AC, Hughes S, Ahmad S, Landau DB, Hawkes DJ. A continuous 4D motion model from multiple respiratory cycles for use in lung radiotherapy. *Med Phys*. 2006;33(9):3348-58. PubMed PMID: 17022231.
66. Mori S, Endo M, Kohno R, Asakura H, Kohno K, Yashiro T, Komatsu S, Kandatsu S, Baba M. Respiratory correlated segment reconstruction algorithm towards four-dimensional radiation therapy using carbon ion beams. *Radiother Oncol*. 2006;80(3):341-8. doi: 10.1016/j.radonc.2006.07.035. PubMed PMID: 16949695.

67. Low DA, White BM, Lee PP, Thomas DH, Gaudio S, Jani SS, Wu X, Lamb JM. A novel CT acquisition and analysis technique for breathing motion modeling. *Phys Med Biol*. 2013;58(11):L31-6. doi: 10.1088/0031-9155/58/11/L31. PubMed PMID: 23640212; PubMed Central PMCID: PMC3685480.
68. Brock KK, Sharpe MB, Dawson LA, Kim SM, Jaffray DA. Accuracy of finite element model-based multi-organ deformable image registration. *Med Phys*. 2005;32(6):1647-59. Epub 2005/07/15. PubMed PMID: 16013724.
69. Hanley JA, McNeil BJ. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology*. 1982;143(1):29-36. PubMed PMID: WOS:A1982NG95400006.
70. Metz CE. ROC analysis in medical imaging: a tutorial review of the literature. *Radiological physics and technology*. 2008;1(1):2-12. doi: 10.1007/s12194-007-0002-1. PubMed PMID: 20821157.
71. Briechle K, Hanebeck UD. Template matching using fast normalized cross correlation. *P Soc Photo-Opt Ins*. 2001;4387:95-102. doi: Doi 10.1117/12.421129. PubMed PMID: WOS:000169645300011.
72. Proschan MA, Waclawiw MA. Practical guidelines for multiplicity adjustment in clinical trials. *Control Clin Trials*. 2000;21(6):527-39. PubMed PMID: 11146147.
73. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology*. 2005;16(1):73-81. Epub 2004/12/23. PubMed PMID: 15613948.
74. Balter JM, Lam KL, McGinn CJ, Lawrence TS, Ten Haken RK. Improvement of CT-based treatment-planning models of abdominal targets using static exhale imaging. *Int J Radiat Oncol Biol Phys*. 1998;41(4):939-43. PubMed PMID: 9652861.
75. Starkschall G, Balter P, Britton K, McAleer MF, Cox JD, Mohan R. Interfractional reproducibility of lung tumor location using various methods of respiratory motion mitigation. *Int J Radiat Oncol Biol Phys*. 2011;79(2):596-601. Epub 2010/07/20. doi: 10.1016/j.ijrobp.2010.03.039. PubMed PMID: 20638189.

76. Nakamura M, Narita Y, Sawada A, Matsugi K, Nakata M, Matsuo Y, Mizowaki T, Hiraoka M. Impact of motion velocity on four-dimensional target volumes: a phantom study. *Med Phys*. 2009;36(5):1610-7. Epub 2009/06/24. PubMed PMID: 19544777.
77. Louie AV, Rodrigues G, Olsthoorn J, Palma D, Yu E, Yaremko B, Ahmad B, Aivas I, Gaede S. Inter-observer and intra-observer reliability for lung cancer target volume delineation in the 4D-CT era. *Radiother Oncol*. 2010;95(2):166-71. Epub 2010/02/04. doi: 10.1016/j.radonc.2009.12.028. PubMed PMID: 20122749.
78. Riegel AC, Chang JY, Vedam SS, Johnson V, Chi PC, Pan T. Cine computed tomography without respiratory surrogate in planning stereotactic radiotherapy for non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys*. 2009;73(2):433-41. doi: 10.1016/j.ijrobp.2008.04.047. PubMed PMID: 18644683.
79. Hunjan S, Starkschall G, Rosen I, Prado K, Tolani N, Balter P. Comparison of breath-hold and free-breathing positions of an external fiducial by analysis of respiratory traces. *J Appl Clin Med Phys*. 2008;9(3):2768. Epub 2008/08/22. PubMed PMID: 18716589.
80. Hunjan S, Starkschall G, Prado K, Dong L, Balter P. Lack of correlation between external fiducial positions and internal tumor positions during breath-hold CT. *Int J Radiat Oncol Biol Phys*. 2010;76(5):1586-91. doi: 10.1016/j.ijrobp.2009.08.008. PubMed PMID: 20133074.
81. McCollough C, Cody D, Edyvean S, Geise R, Gould B, Keat N, Huda W, Judy P, Kalender W, McNitt-Gray M, Morin R, Payne T, Stern S, Rothenberg L, Shrimpton P, Timmer J, Wilson C. The Measurement, Reporting, and Management of Radiation Dose in CT. *AAPM* 2008:34
82. Kalender WA. Dose in x-ray computed tomography. *Phys Med Biol*. 2014;59(3):R129-50. doi: 10.1088/0031-9155/59/3/R129. PubMed PMID: 24434792.
83. Shrimpton P. Assessment of Patient Dose in CT. National Radiological Protection Board Report 2004. p. 36.
84. Shrimpton PC, Hillier MC, Lewis MA, Dunn M. National survey of doses from CT in the UK: 2003. *The British journal of radiology*. 2006;79(948):968-80. doi: 10.1259/bjr/93277434. PubMed PMID: 17213302.

85. Li R, Lewis JH, Cervino LI, Jiang SB. 4D CT sorting based on patient internal anatomy. *Phys Med Biol*. 2009;54(15):4821-33. doi: 10.1088/0031-9155/54/15/012. PubMed PMID: 19622855.
86. Eck K, Bredno J, Stehle T. Absolute alignment of breathing states using image similarity derivatives. *Medical Imaging 2005: Visualization, Image-Guided Procedures, and Display, Pts 1 and 2*. 2005;5744:79-86. doi: Doi 10.1117/12.595313. PubMed PMID: WOS:000229312400010.
87. Castillo SJ, Castillo R, Balter P, Pan T, Ibbott G, Hobbs B, Yuan Y, Guerrero T. Assessment of a Quantitative Metric for 4D CT Artifact Evaluation by Observer Consensus *Journal of Applied Clinical Medical Physics* 2014.
88. Bland JM, Altman DG. Measurement error. *BMJ*. 1996;313(7059):744. Epub 1996/09/21. PubMed PMID: 8819450; PubMed Central PMCID: PMC2352101.
89. Bland JM, Altman DG. Measurement error and correlation coefficients. *BMJ*. 1996;313(7048):41-2. Epub 1996/07/06. PubMed PMID: 8664775; PubMed Central PMCID: PMC2351452.
90. Busha BF, Hage E, Hofmann C. Gender and breathing route modulate cardio-respiratory variability in humans. *Respir Physiol Neurobiol*. 2009;166(2):87-94. doi: 10.1016/j.resp.2009.02.008. PubMed PMID: 19429524.
91. Romei M, Mauro AL, D'Angelo MG, Turconi AC, Bresolin N, Pedotti A, Aliverti A. Effects of gender and posture on thoraco-abdominal kinematics during quiet breathing in healthy adults. *Respir Physiol Neurobiol*. 2010;172(3):184-91. doi: 10.1016/j.resp.2010.05.018. PubMed PMID: 20510388.
92. Ragnarsdottir M, Kristinsdottir EK. Breathing movements and breathing patterns among healthy men and women 20-69 years of age. Reference values. *Respiration; international review of thoracic diseases*. 2006;73(1):48-54. doi: 10.1159/000087456. PubMed PMID: 16106113.

Vita

Sarah Joy Castillo was born October 14, 1987 at Offutt Air Force Base in Nebraska. She completed her physics B.S. degree from the University of Florida in 2009, and her M.S. degree from the University of Texas Houston Health Science Center in 2011.

Permanent Address:

None