APPENDIX: Concordance of physical and simulation lesions for radiomics quantification

A.I. RATIONALE

Simulation studies are useful and valuable for studying the effects of Computed Tomography (CT) imaging protocols on the ability to measure quantitative features accurately and reliably from images. While simulation studies are efficient and more easily controlled than physical studies, there is a need to show agreement between simulation and physical studies for simulation studies to be trusted. Therefore, in this study, we performed an analysis to compare the bias and variability of radiomics features extracted from physical phantoms imaged on real commercial CT scanners and those extracted from digital phantoms imaged via simulations designed to replicate the physical scanner.

A.II. METHODS

A parallel analysis of the bias and variability of morphology radiomics features was performed for (1) a physical lesion model scanned with a real CT scanner and (2) a matching computational lesion model virtually scanned with a CT simulation platform. The bias and variability of radiomics were compared between the physical and simulation studies to assess the agreement between simulation and physical studies.

Physical Data

A 32 cm cylindrical physical phantom containing anthropomorphic lesion models was previously 3D printed.[1] The phantom was scanned four repeated times with a commercial CT scanner (Siemens Flash) with 120 kVp, 0.5 s rotation time, pitch of 1, and 128 × 0.6 mm collimation. The $CTDI_{vol}$ was 22 mGy and the images were reconstructed with 0.4688 mm pixel spacing, 0.6 mm slice thickness, and two filtered back projection reconstruction kernels (B31f

and B45f). A lesion model was isolated in the phantom with intensity around 150 HU and with volume of 2145 mm^3. For each repeated image, the lesion model was segmented using a MATLAB implementation of an active contour segmentation approach[2]. The segmentation resulted in a 3-dimensional logical mask which was re-sampled to isotropic voxel size of 0.5 mm. The resampled segmentation mask was used to calculate a series of morphology radiomics features described in Hoye et al.[3]

Simulated Data

A computational ground truth mesh-based representation of the lesion model used to create the 3D printed phantom was used for the simulation study. The computational lesion model was created with the same properties as the corresponding physical model including the size, shape, and intensity properties. Simulated images of the lesion models were created by applying image system blur using the corresponding measured Task Transfer Function (TTF)[4] and Noise Power Spectrum (NPS) for each reconstruction method. The method of generating the simulated images has been described previously.[3] In the present study, the in-plane TTF and NPS curves were obtained by measurements from the physical cylindrical phantom as described by Robins et al.[1] The noise magnitude for each simulated image was implemented to match the noise magnitude for the corresponding physical phantom images. The z-dimension TTF was implemented using an idealized theoretical model described in Hoye et al.[3] For the simulation, the noise between image slices was assumed uncorrelated.

The simulated images were generated 4 repeated times for each reconstruction kernel by using different random numbers in the noise addition process. For each repeated simulated image, the computational lesion model was segmented using a MATLAB implementation of an active contour segmentation approach.[2] The segmentation resulted in a 3-dimensional logical

mask which was re-sampled to isotropic voxel size of 0.5 mm. The resampled segmentation mask was used to calculate a series of morphology radiomics features described in Hoye et al.[3].

Analysis

The morphology features for the physical and simulation arms of the study were compared in terms of the bias and variability of bias. The bias was calculated by comparing the measured radiomics features from the physical and simulated images to the ground truth radiomics features calculated from the ground truth computational model that was used for both the basis of the simulated and physical studies. The relative bias (B) was calculated as

$$\mathrm{B} = 100 \times \left(\frac{F_M - F_0}{F_0}\right), \quad (1)$$

where F_M is the measured radiomics feature from either the physical or simulated image and F_0 is the ground truth radiomics feature value. The variability of the bias was calculated as the standard deviation of the bias values among the 4 repeats for each reconstruction kernel (B31f and B45f).

A.III. RESULTS

The images of the lesion model for real and simulated data showed good visual agreement in noise, blur, and contrast properties (Figure H). The bias and standard deviation of the bias results showed close agreement between physical and simulated data for most features among the two reconstruction kernels (Figure G). The average difference in the bias between physical and simulation was 3.9% for the B31f kernel and 2.5% for the B45f kernel. The differences in bias between the physical and simulation studies were less than 5% for most features and were less than 21% for all features. The average difference in the variability was

0.3% for the B31f kernel and 2% for the B45f kernel. The difference in variability was less than 5% for all features except for one case where the difference was 14% (Asphericity, B45f).

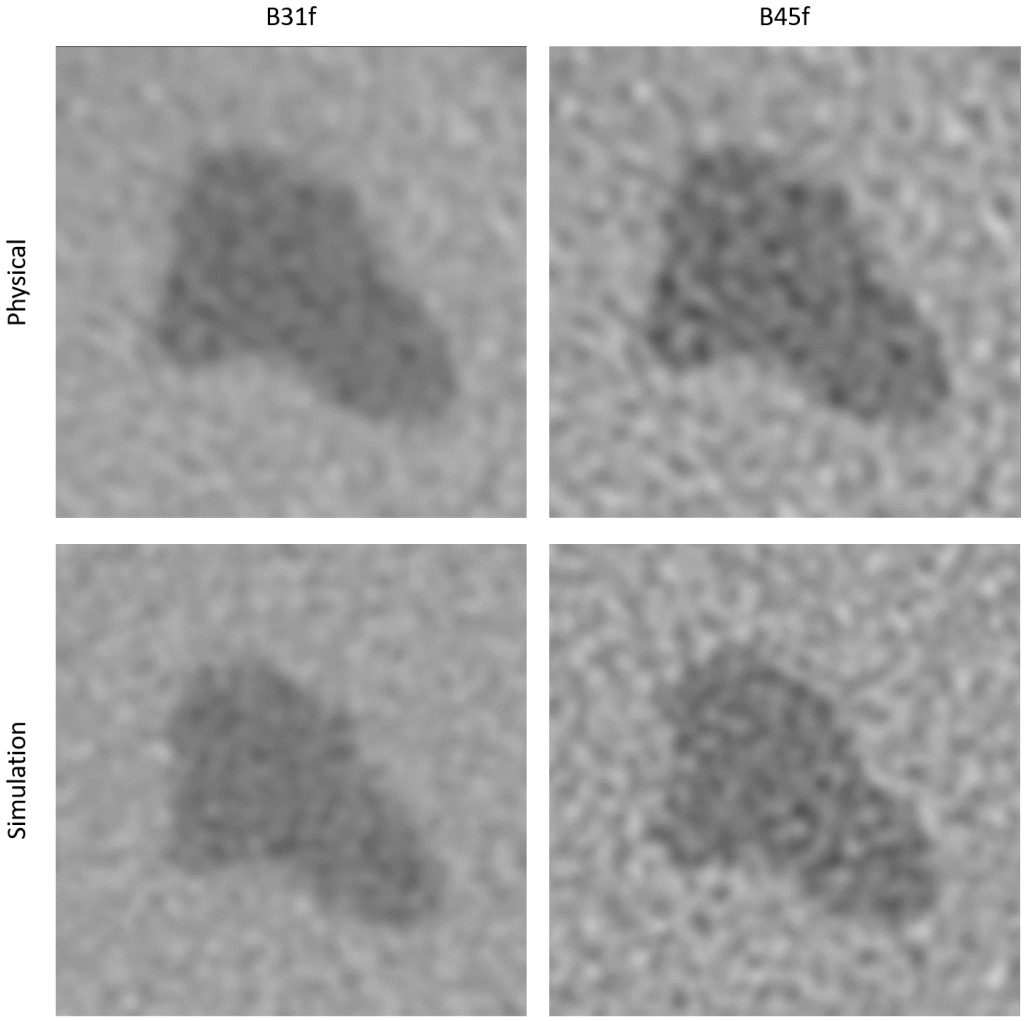

Figure H. Example images of the region of interest around the lesion model for both the physical and simulation studies for the B31f and B45f reconstruction kernels. All images are displayed with window width of 400 and window level of 100. To displaying the images, axial slices were matched as close as possible between physical and simulation to the same location within the lesion. However, the lesion axial slices could not be exactly aligned due to differences in the z sampling position of the lesion model relative to the image slice between physical and simulation.

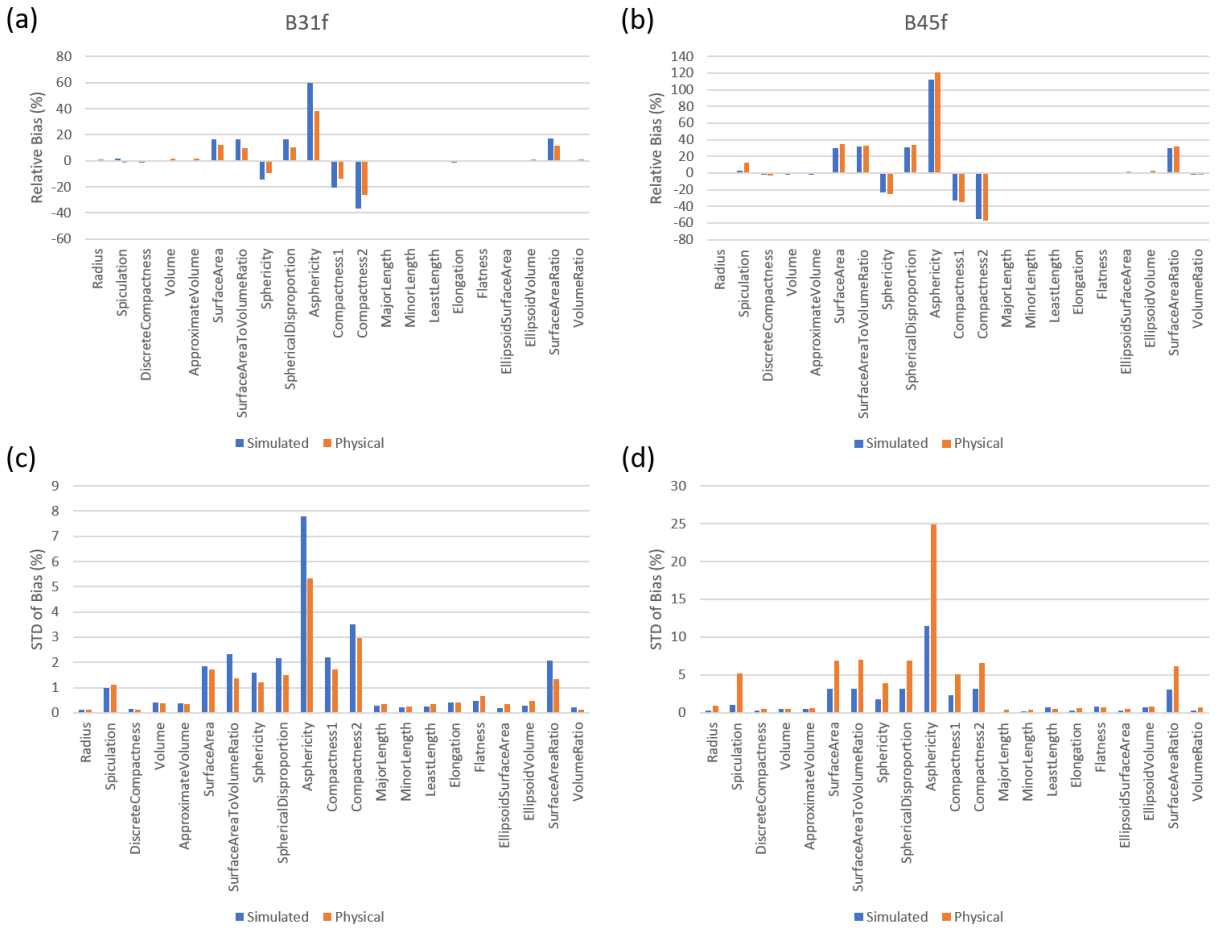

Figure G. The average relative bias (%) among the 4 repeated images for the physical and simulated radiomics features for the B31f kernel (a) and the B45f kernel (b). The standard deviation of the bias (%) among the 4 repeated images for the physical and simulation radiomics features for the B31f kernel (c) and the B45f kernel (d). The differences in bias were less than 5% for most features and less than 21% for all features. The differences in variability were less than 5% for most features and less than 14% for all features.

A.IV. DISCUSSION AND CONCLUSION

This study completed its aim of comparing the bias and variability between real and simulated images of anthropomorphic lesion models. The simulated and physical images had good visual agreement and the bias and variability values were within 5% for most cases and were less than 21% for all cases. The features that showed larger differences in the bias and

variability values between physical and simulation were features that are based on equations that utilize surface area and volume measurements. As a result, we expect that small differences in the volume and surface area properties are propagated into larger differences for these features. There were also some inherent variabilities in the simulation and physical setups that may have resulted in differences in the bias and variability. For example, one source of variability was how the z-sampling lined up with the anatomical locations in the physical and simulation studies. While we attempted to line it up the physical and simulation data as closely as possible, some slight differences remained. Given that the z-sampling locations were fixed among the repeats for both the physical and simulation studies, this variability represents a potential source of systematic bias between the two arms of the study. Future work should seek to characterize the variability induced by different z-sampling.

In conclusion, for the reconstruction protocols tested, the results show the disagreement between real and physical bias and variability measurements to be less than 21% overall and less than 5% for most features.

References

1. Robins M, Solomon J, Richards T, Samei E. 3D task-transfer function representation of the signal transfer properties of low-contrast lesions in FBP-and iterative-reconstructed CT. *Medical physics.* 2018;45(11):4977-4985.
2. Whitaker RT. A level-set approach to 3D reconstruction from range data. *International journal of computer vision.* 1998;29(3):203-231.
3. Hoye J, Solomon J, Sauer TJ, Robins M, Samei E. Systematic analysis of bias and variability of morphologic features for lung lesions in computed tomography. *JMIOBU.* 2019;6(1):013504.
4. Richard S, Husarik DB, Yadava G, Murphy SN, Samei E. Towards task-based assessment of CT performance: System and object MTF across different reconstruction algorithms. *Medical physics.* 2012;39(7Part1):4115-4122.