Comparison of PET Contouring Methods in Patients with Early Stage Resected Non-Small Cell Lung Cancer (NSCLC): A Pathologic-Imaging Correlation

M. Werner-Wasik[1], P. Kang[2], W. Choi[2], N. Ohri[3], P. Faulhaber[4], D. Nelson[5], A. Nelson[5], J. Piper[5], Z. Shen[5], S. Pirozzi[5]

[1]Thomas Jefferson University Hospital, Philadelphia, PA, [2]Beth Israel Medical Center, New York City, NY,
[3]Montefiore Medical Center, Albert Einstein College of Medicine, New York City, NY, [4]University Hospitals Case Medical Center, Cleveland, OH,
[5]MIM Software Inc, Cleveland, OH

Purpose

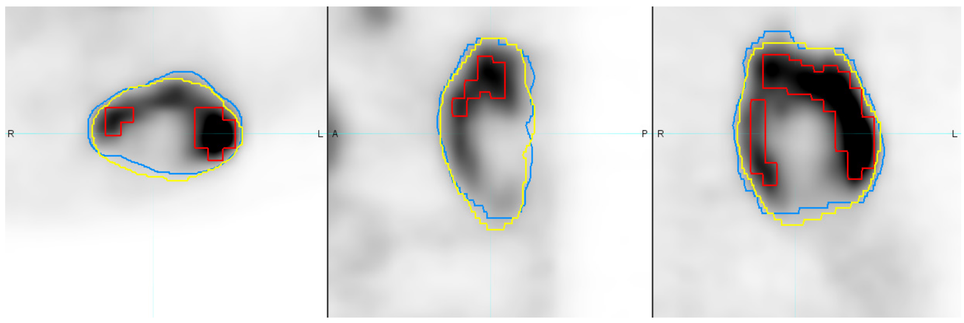
The best way of contouring lung tumors on PET scans is not well known, highlighting the need for an accurate and consistent method for segmenting PET target volumes. In a previous work we demonstrated a gradient-based PET segmentation method (Gradient), was the most accurate technique with the smallest interobserver variability for Monte Carlo simulated digital NSCLC PET phantoms[1]. While offering encouraging results, pathological validation of this Gradient technique is still lacking in NSCLC. The purpose of this study was to compare the contouring techniques used previously in the PET phantom using actual FDG PET scans from patients (pts) with resected early stage NSCLC.

Methods and Materials

Anonymized preoperative FDG-PET CT scan images of pts with early stage NSCLC who underwent lobectomy or sublobar resection between 2006 and 2009 were acquired from one institution's hospital database after IRB approval. PET scans were uploaded to a cloud database, to which 5 experienced observers from different institutions (3 radiation oncologists and 2 nuclear medicine physicians) connected remotely and contoured each lesion using Manual, Gradient, and Threshold contouring methods (from 25% to 50% of max SUV at 5% increments). Observers were asked to wait 24 hours between each method. The longest PET-derived tumor diameter was measured for each contour and compared to the surgical tumor largest diameter obtained from pathologic records of the dissected tumor specimens. Inter-observer variability was calculated using Levene's test. Bland–Altman analysis of differences vs. means of paired values was used to search for trends and systematic errors.

Figure 1

Example Patient Case: Heterogeneous NSCLC Tumor



Manual ■ Threshold (50%) ■ Gradient ■

FDG-PET scan of a NSCLC tumor with Manual, 50% Threshold, and Gradient segmentations. Hetergeneous tumor with low activity illustrating a limitation of threshold methods while Manual and Gradient maintained accuracy.

Reference

1 Werner-Wasik M, Nelson A, Choi W, et al. What is the best way to contour lung tumors on PET scans? Multiobserver validation of a a gradient-based method using a NSCLC Digital PET Phantom. IJROBP 82; pp1164-1171, 2012.

Author Disclosure

M. Werner-Wasik: None. P. Kang: None. W. Choi: None. N. Ohri: None. P. Faulhaber: None. D. Nelson: A. Employee; MIM Software. A. Nelson: A. Employee; MIM Software. J. Piper: A. Employee; MIM Software. Z. Shen: A. Employee; MIM Software. S. Pirozzi: A. Employee; MIM Software.

Results

Images of 46 pts were used for contouring. Pathologic tumor diameters averaged 2.55 +/- 1.42 cm [0.7cm, 6.5cm]. For all tumors, Manual and Gradient techniques had the best correlations to pathology (r = 0.79 vs. 0.76 respectively). Thresholds had lower correlations ranging from 0.61 (35% Threshold) to 0.66 (50% Threshold). Manual and Gradient techniques also had the smallest absolute differences in tumor diameter (0.76 vs. 0.84) when compared to pathologic tumor diameter, while Thresholds ranged from 0.86 (50% Threshold) to 1.29 (25% Threshold). Gradient and 50% Threshold (the most accurate threshold method) were found to have less interobserver variability than Manual technique ($p < 0.01$), but the difference between Gradient and 50% Threshold was not significant.

Figure 2

Bland-Altman plots comparing pathological largest measured diameters to manual and automated segmentation techniques. Gradient gave the tightest confidence interval over all other automatic methods.

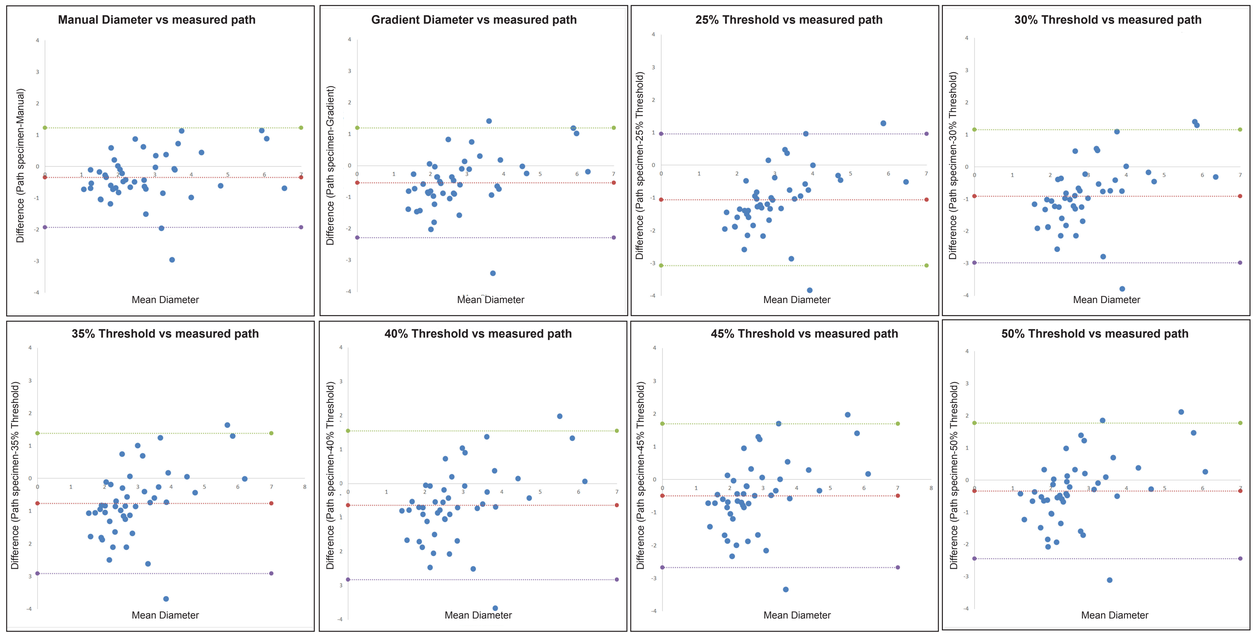


Table 1

PET Segmentation Methods versus Pathology Max Diameter

	Manual	Gradient	25% Thresh	30% Thresh	35% Thresh	40% Thresh	45% Thresh	50% Thresh
Average Difference (cm)	-0.356	-0.544	-1.057	-0.908	-0.763	-0.638	-0.489	-0.334
Average Absolute Difference (cm)	0.763	0.841	1.290	1.174	1.092	1.016	0.955	0.856
Correlation	0.791	0.762	0.628	0.628	0.611	0.609	0.616	0.656

Conclusion

Gradient and Manual contouring both correlated well with pathology longest diameter. Gradient, however, was more consistent between observers than Manual. The 50% Threshold method was also more consistent than Manual, however, had significantly lower correlation with pathology than both Gradient and Manual.