Int. J. Radiation Oncology Biol. Phys., Vol. 82, No. 3, pp. 1164–1171, 2012

0360-3016/$ - see front matter

doi:10.1016/j.ijrobp.2010.12.055

CLINICAL INVESTIGATION **Thoracic Cancer**

WHAT IS THE BEST WAY TO CONTOUR LUNG TUMORS ON PET SCANS? MULTIOBSERVER VALIDATION OF A GRADIENT-BASED METHOD USING A NSCLC DIGITAL PET PHANTOM

Maria Werner-Wasik, M.D.,* Arden D. Nelson, Ph.D.,† Walter Choi, M.D.,§ Yoshio Arai, M.D.,‡ Peter F. Faulhaber, M.D.,¶ Patrick Kang, M.D.,‖ Fabio D. Almeida, M.D.,†† Ying Xiao, Ph.D.,* Nitin Ohri, M.D.,* Kristin D. Brockway, B.S.,† Jonathan W. Piper, B.S.,† and Aaron S. Nelson, M.D.†

*Department of Radiation Oncology, Thomas Jefferson University Hospital, Philadelphia, PA; †MIM Software Inc., Cleveland, OH; ‡Departments of Radiation Oncology and ‖Radiology, Beth Israel Medical Center, New York, NY; §Department of Radiation Oncology, UPMC Health Systems, Pittsburgh, PA; ¶University Hospitals Case Medical Center, Cleveland, OH; and ††Division of Nuclear Medicine, University of Arizona Health Sciences Center, Tucson, AZ

Purpose: To evaluate the accuracy and consistency of a gradient-based positron emission tomography (PET) segmentation method, GRADIENT, compared with manual (MANUAL) and constant threshold (THRESHOLD) methods.
Methods and Materials: Contouring accuracy was evaluated with sphere phantoms and clinically realistic Monte Carlo PET phantoms of the thorax. The sphere phantoms were 10–37 mm in diameter and were acquired at five institutions emulating clinical conditions. One institution also acquired a sphere phantom with multiple source-to-background ratios of 2:1, 5:1, 10:1, 20:1, and 70:1. One observer segmented (contoured) each sphere with GRADIENT and THRESHOLD from 25% to 50% at 5% increments. Subsequently, seven physicians segmented 31 lesions (7–264 mL) from 25 digital thorax phantoms using GRADIENT, THRESHOLD, and MANUAL.
Results: For spheres <20 mm in diameter, GRADIENT was the most accurate with a mean absolute % error in diameter of 8.15% (10.2% SD) compared with 49.2% (51.1% SD) for 45% THRESHOLD ($p < 0.005$). For larger spheres, the methods were statistically equivalent. For varying source-to-background ratios, GRADIENT was the most accurate for spheres >20 mm ($p < 0.065$) and <20 mm ($p < 0.015$). For digital thorax phantoms, GRADIENT was the most accurate ($p < 0.01$), with a mean absolute % error in volume of 10.99% (11.9% SD), followed by 25% THRESHOLD at 17.5% (29.4% SD), and MANUAL at 19.5% (17.2% SD). GRADIENT had the least systematic bias, with a mean % error in volume of –0.05% (16.2% SD) compared with 25% THRESHOLD at –2.1% (34.2% SD) and MANUAL at –16.3% (20.2% SD; p value <0.01). Interobserver variability was reduced using GRADIENT compared with both 25% THRESHOLD and MANUAL (p value <0.01, Levene's test).
Conclusion: GRADIENT was the most accurate and consistent technique for target volume contouring. GRADIENT was also the most robust for varying imaging conditions. GRADIENT has the potential to play an important role for tumor delineation in radiation therapy planning and response assessment. © 2012 Elsevier Inc.

PET scan, Lung cancer, Tumor segmentation, Radiation therapy planning.

INTRODUCTION

Fluorodeoxyglucose positron emission tomography (FDG-PET) scans are used in lung cancer management for the initial staging (1), radiation therapy (RT) planning (2), and the evaluation of tumor response to therapy (3, 4). Advances in radiation therapy technology have improved the ability to deliver highly conformal therapy for smaller tumors and have increased the need for accurate and consistent definition of tumor boundaries. It has been demonstrated that applying PET to RT planning changes gross target volume (GTV) in more than 50% of non–small cell lung cancer (NSCLC) cases (2) and is particularly valuable in patients whose tumors blend with atelectasis in computed tomography (CT) image volumes. Great interobserver variability has been reported in CT definition of GTV in lung cancer (5), indicating the limitations of CT for tumor definition. Therefore, there is great interest in lowering this variability, possibly with application of PET

Reprint requests to: Maria Werner-Wasik, M.D., Department of Radiation Oncology, Thomas Jefferson University Hospital, 111 South 11th Street, Bodine G301E, Philadelphia, PA 19107. Tel: (215) 955-7679; Fax: (215) (955)-0412; E-mail: Maria.Werner-wasik@jeffersonhospital.org

Conflicts of interest: none.

Acknowledgments—We thank Micahlis Aristophanous, Bill Penney, and Charles Pelizzari for supplying the Monte Carlo thorax phantoms used in this study (see reference 12).

Received June 18, 2010, and in revised form Nov 24, 2010. Accepted for publication Dec 19, 2010.

images for tumor delineation. Before PET can be widely applied for this purpose, standards must be established for the contouring technique.

There is currently no consensus as to the optimal technique for delineating (segmenting) PET target volumes. Various approaches are used, including the following:

1. Manual contouring (MANUAL), in which the physician determines the tumor outline on the basis of visual perception of the tumor border.
2. Threshold methods, which define the tumor border within a region-of-interest placed over the tumor by including all tissue with activity greater then a defined level. Absolute thresholds define the tumor border on the basis of a minimum SUV level. Suggested standardized uptake value levels have included 2.0 (5), 2.5 (6), or 3.0 ± 1.6 from a recent study looking for the absolute threshold level that produced volumes most similar to pathology measurements for nine NSCLC patients (7). Percent constant threshold methods (THRESHOLD) define the tumor border on the basis of a percentage of the maximum activity within the tumor. All tissue with activity greater than that percentage is included within the tumor volume. The impact of lesion size and source-to-background ratio on volumes obtained with constant threshold methods has been reported previously (1, 8–10). A recent study demonstrated that to obtain image-derived volumes equal to pathology volumes in nine NSCLC patients, constant thresholds levels between 20% and 42% of maximum were required (7). Adaptive threshold methods use parameters such as tumor size and the ratio of tumor to background levels to define the threshold level (8). Currently, there is no consensus as to the appropriate threshold method or best threshold levels for tumor segmentation. This variability is one factor limiting use of PET for tumor definition in radiation oncology. Most clinicians continue to rely on the CT-derived volume as the gold standard for GTV contouring and use PET as an ancillary tool, mostly to prevent omitting hypermetabolic areas from patient's GTV or to identify the interface between tumor and atelectasis.
3. Gradient edge detection identifies tumor on the basis of a change in count levels at the tumor border. One proposed method requires, in the following order, a denoising tool, a deblurring tool, a gradient estimator, and a watershed transform (11). This method is sensitive to voxel size, varying image resolution, and noise, which requires adjusting one or more of these tools and making it less realistic for routine clinical use. The gradient method evaluated in this article, GRADIENT (MIM Software, Cleveland, OH), calculates spatial derivatives along tumor radii then defines the tumor edge on the basis of derivative levels and continuity of the tumor edge.

Our goal was to evaluate the accuracy, bias, and consistency of GRADIENT compared with traditional manual and percent threshold contouring methods. In this article, we first used the experimental sphere phantoms to evaluate the impact of various PET cameras, sphere sizes, reconstruction methods, and source-to-background ratios on border detection with both THRESHOLD and GRADIENT. Subsequently, to emulate clinical reality more closely, we evaluated and compared three methods of PET tumor contouring: MANUAL, THRESHOLD, and GRADIENT. We used Monte Carlo PET thorax phantoms (12), which have been designed to simulate both lung tumors and mediastinal lymph node metastases. Because the true volumes of these simulated tumors and lymph nodes are known, they serve as the gold standard for the volumes contoured by the physicians.

METHODS AND MATERIALS

Contouring methods

MANUAL: Each observer used a manual contouring tool of their choice (pen, 2D, or 3D paintbrush) provided in MIM (MIM Software) to delineate the structure of interest by visually outlining the boundaries. Five observers used both 3D and 2D brushes, one used 3D only, and one used pen only. The structure could be contoured in any cross-section and viewed in either a single slice or a splash page of contiguous slices. Each observer was able to adjust image contrast levels according to his or her own preference to allow for optimal visualization of the structure.

THRESHOLD: The THRESHOLD contouring method relies on including all voxels that are greater than a defined percent of the maximum voxel within an operator-defined sphere. Cross-sectional circles are displayed in all three projections (axial, sagittal, and coronal) as the operator defines the sphere size and location to ensure three-dimensional coverage of the structure of interest. The structure could be contoured in any cross-section and viewed in either a single slice or a splash page of contiguous slices. Each observer adjusted image contrast levels according to his or her own preference to allow for optimal visualization of the structure.

GRADIENT: The gradient method relies on an operator-defined starting point near the center of the lesion. As the operator drags out from the center of the lesion, six axes extend out, providing visual feedback for the starting point of gradient segmentation. Spatial gradients are calculated along each axis interactively, and the length of an axis is restricted when a large spatial gradient is detected along that axis. The six axes define an ellipsoid that is then used as an initial bounding region for gradient detection. The observers in the study were instructed to begin by selecting the image slice in which they identified the tumor to appear largest. The observer was then instructed to localize at a point near the center of the lesion in this slice and drag from that point until the six axes approximated the boundaries of the lesion (Fig. 1). After releasing the mouse button, the edges of the structure were automatically calculated and outlined. For very irregularly shaped structures, which are not well defined by the six axes, observers were instructed to use the gradient method one or more times to add to the initial contour by dragging out from a point near the center of the omitted region. Operators added regions until they were visually satisfied that the entire structure was included in the contour.

Characteristics of sphere phantoms

PET scans were acquired for commercially available sphere phantoms with five PET scanners at five institutions (Table 1). The institutions were instructed to emulate clinical acquisition and reconstruction methods used at that institution. All institutions were instructed to acquire the phantom with source to background