Medical Dosimetry 41 (2016) 344–350

Medical Dosimetry

journal homepage: www.meddos.org

Anatomical contouring variability in thoracic organs at risk

Ross McCall, M.S., CMD,* Grayden MacLennan, M.B.A., M.S.M., M.S., CMD,* Matthew Taylor, M.S.,* Nishele Lenards, M.S., CMD, FAAMD,* Benjamin E. Nelms, Ph.D.,† Matthew Koshy, M.D.,‡ Jeffrey Lemons, M.D.,‡ and Ashley Hunzeker, M.S., CMD*

*Medical Dosimetry Program, University of Wisconsin, La Crosse, WI; †Canis Lupus LLC, Madison, WI; and ‡Department of Radiation and Cellular Oncology, University of Chicago, Chicago, IL

ARTICLE INFO

Article history:
Received 12 February 2016
Received in revised form
11 July 2016
Accepted 9 August 2016

Keywords:
Contouring
Organs at risk
Contour variability
Thorax
Heart
Esophagus
Spine

ABSTRACT

The purpose of this study was to determine whether contouring thoracic organs at risk was consistent among medical dosimetrists and to identify how trends in dosimetrist's education and experience affected contouring accuracy. Qualitative and quantitative methods were used to contextualize the raw data that were obtained. A total of 3 different computed tomography (CT) data sets were provided to medical dosimetrists ($N = 13$) across 5 different institutions. The medical dosimetrists were directed to contour the lungs, heart, spinal cord, and esophagus. The medical dosimetrists were instructed to contour in line with their institutional standards and were allowed to use any contouring tool or technique that they would traditionally use. The contours from each medical dosimetrist were evaluated against "gold standard" contours drawn and validated by 2 radiation oncology physicians. The dosimetrist-derived contours were evaluated against the gold standard using both a Dice coefficient method and a penalty-based metric scoring system. A short survey was also completed by each medical dosimetrist to evaluate their individual contouring experience. There was no significant variation in the contouring consistency of the lungs and spinal cord. Intradosimetrist contouring was consistent for those who contoured the esophagus and heart correctly; however, medical dosimetrists with a poor metric score showed erratic and inconsistent methods of contouring.

Introduction

Accurate delineation of organs at risk (OR) on patient imaging is an essential step in the protection of sensitive anatomy when designing radiation treatment plans. The boundaries of target volumes and OR are established in the treatment planning system by drawing regions of interest (ROI) on the patient image set. By designating a section of the image as part of a particular organ or target (or both), useful statistics can be extracted such as structure size and radiation dose coverage across the volume. A dose-volume histogram analysis can show the absolute or relative volume of any mapped structure that receives any given amount of dose. The dose-volume histogram is a commonly used tool for evaluation of plan quality based on dose coverage of targets and sparing of normal tissues. Therefore, it is paramount that the contouring of OR be done accurately by the medical dosimetrist.

During radiation therapy planning for thoracic treatments, the dose received by the lungs, heart, esophagus, and spinal cord must be evaluated. The potential harm from irradiating these organs must be carefully considered when designing the treatment plan. The radiation dose tolerances and physiological consequences of radiation injury have been studied extensively for each of these organs.[1-4] Organs with functional subunits arranged in a serial fashion such as the spinal cord and esophagus are especially vulnerable owing to the potential failure of the entire organ when a single subunit is exposed to a high dose of radiation.[5] Serial organs are, therefore, especially sensitive to errors in contouring when near high-dose regions.

As imaging, planning, and delivery technologies and techniques have evolved, it has become possible to decrease dose to OR while maintaining or escalating the dose to target volumes. This is accomplished by using tightly conforming dose planning techniques such as intensity-modulated radiation therapy or volumetric-modulated arc therapy.[6,7] These techniques can be combined with immobilization techniques and image-guided radiotherapy technology to safely deliver dose plans with steep dose gradients near critical OR. When high-dose regions are placed closer to critical OR, it becomes increasingly important to ensure that they are contoured correctly.[8]

Reprint requests to Ross McCall, M.S., CMD, 8 E. Randolph St., Unit 1206, Chicago, IL 60601.
E-mail: rmccall86@gmail.com

http://dx.doi.org/10.1016/j.meddos.2016.08.004
0958-3947/

Contouring remains one of the largest sources of uncertainty in the radiation planning and delivery process.[9] Inaccuracies in the definition of ROI would create inaccuracies in dose evaluation resulting in the potential for damage to sensitive OR. The quantitative analysis used in this study computed contouring accuracy by using a score that involved not just absolute degree of overlap but also a bonus and penalty system based on distance to agreement as described by Nelms *et al.*[10] This scoring system applied penalties that escalated with distance to agreement, allowing a more nuanced evaluation. Furthermore, evaluation of each medical dosimetrist's experience and education was performed to identify trends in contouring accuracy as well as identify opportunities for improvements in training.

Study of contouring accuracy and precision has been done on several regions of the body.[10,11] A literature review was performed to identify areas of opportunity in which data were sparse. Some research was found to have been done on OR delineation in the head and neck region as well as the thorax; however, most contouring research focused on target delineation rather than OR that were contoured by medical dosimetrists. In addition to this, no available research was found that studied intradosimetrist variation or identifiable trends in medical dosimetrist background *vs* their performance.

Methods and Materials

Overall, 3 CT data sets of patients with conventional thoracic anatomy were obtained from previous total-marrow irradiation research. The data sets were helical CT scans sampled at 3-mm slice thickness. The image sets were cropped from the mandible to just below the xyphoid process or longer as needed to cover the entire extent of the esophagus and lungs.

A “gold standard” contour set consisting of the left lung, right lung, heart, esophagus, and spinal cord was created for each data set by a radiation oncologist according to the Radiation Therapy Oncology Group (RTOG) 1106 Lung Atlas.[12] The gold standard sets were then peer reviewed by a fellow radiation oncologist for accuracy. In total, 13 medical dosimetrists across 5 institutions were then asked to contour the same 5 structures with the same time, care, and precision that would be used in daily clinical practice. The medical dosimetrists were given no restrictions regarding which contouring software package or which drawing tools they were allowed to use.

Following the contouring, the medical dosimetrists were directed to complete a survey gathering information about training in medical dosimetry, certification status, years of experience, and any continuing education related to contouring that was completed. The medical dosimetrists were also asked about their familiarity with RTOG guidelines for contouring the structures.

A statistical analysis of each medical dosimetrist's contours was performed with a custom-coded variant of the StructSure (Standard Imaging) contouring quality assurance package. Dice similarity coefficients and StructSure's proprietary scoring system were both used. The Dice similarity coefficients was computed as the intersection volume between the tested set and the “gold standard” set divided by the union volume of both sets.

$$\text{Dice coefficient} = \frac{\text{ROI}_{\text{test}}\text{IROI}_{\text{gold}}}{\text{ROI}_{\text{test}}\text{UROI}_{\text{gold}}}$$

StructSure's proprietary method provided scoring of contours based on a linear distance to agreement algorithm that assigned positive scores to voxels that overlap a known gold standard contour, and negative scores to voxels that did not overlap, with the total score then normalized to the total number of gold voxels.[10] Penalties varied with distance so that minor deviations earned small penalties, whereas larger distances to agreement earned greater penalties. To differentiate variations owing to misinterpretation of the CT images from variations because of unsteadiness of hands or contour resolution limitations, the penalty equation was adjusted to further emphasize the penalties on large deviations, whereas ignoring small deviations completely. Similar to the method used by Nelms *et al.*,[10] a discontinuous penalty equation was created, granting 2 mm of “forgiveness” margin, followed by increasing penalties beyond that distance. Subtle edge roughness, therefore, created no penalty in scoring, whereas contouring the wrong piece of anatomy did create a penalty.

The axial contours are input into a “voxelization” routine that creates a high-resolution volume in space. The volume is composed of small 3D pixels, called voxels, whose edge dimensions can be specified in the software. The default dimension of 0.5 mm was used. The “distance-to-agreement” function of the StructSure software used a 0.50/mm penalty for each voxel that was contoured in error (*i.e.*, not within the forgiveness region). The distance calculated for any error voxel from the ROI is calculated by the shortest 3D distance from the tested volume to the gold volume's surface. A score was generated in the 0 to 100 range. A perfect agreement between the gold contour and the medical dosimetrist contour would earn a 100 score.

Contouring results were evaluated qualitatively to identify commonalities in contouring mistakes that could be highlighted for medical dosimetrist education. These were obvious errors that were often associated with RTOG guidelines in which the medical dosimetrist did not properly contour the superior/inferior edges of an OR or wrongly included an errant piece of anatomy. Finally, to provide context for the significance of contouring errors, a mock plan was generated using the gold standard contours on a single data set. An apical planning target volume (PTV) that would mimic a real-world lung lesion was contoured. The mock PTV was placed adjacent to the esophagus in the right lung. A minimum distance of 4 mm was placed between the PTV and the esophagus. The PTV was assigned a Hounsfield unit (HU) of 0. A volumetric-modulated arc therapy technique was chosen using two 360° arcs, and coverage was normalized so that 100% of the prescription dose was covering 95% of the volume. Once an adequate plan was developed, each medical dosimetrist's structure set was imported into the treatment plan, and the maximum and mean doses to each OR were calculated.

Results

Quantitative analysis

A quantitative analysis demonstrated the spinal cord and both lungs were contoured most accurately, whereas the heart and esophagus had the most variation (Table 1). There was no statistically significant correlation identified when contouring accuracy was compared against years of experience, attendance of a formal medical dosimetry training program, or self-reported familiarity

Table 1
Aggregated quantitative finding for all structures ($n = 39$ contours per OR)

	Mean metric score	Standard deviation	Mean Dice coefficient	Standard deviation
Spinal cord	97.83	3.35	0.86	0.02
Left lung	99.13	0.49	0.96	0.01
Right lung	97.61	1.03	0.96	0.01
Heart	90.38	5.92	0.91	0.03
Esophagus	74.40	11.86	0.74	0.03