Contouring variations and the role of atlas in non-small-cell lung cancer radiotherapy: analysis of a multi-institutional pre-clinical trial planning study

Yunfeng Cui, PhD[1], **Wenzhou Chen, PhD**[2], **Feng-Ming (Spring) Kong, MD, PhD**[3], **Lindsey A. Olsen, MSc**[4], **Ronald E. Beatty, MEng**[5], **Peter G. Maxim, PhD**[6], **Timothy Ritter, PhD**[7], **Jason W. Sohn, PhD**[8], **Jane Higgins, BSc**[9], **James M. Galvin, DSc**[2,10], and **Ying Xiao, PhD**[2,10]

[1]Department of Radiation Oncology, Duke University Medical Center, Durham, NC

[2]Department of Radiation Oncology, Jefferson Medical College of Thomas Jefferson University, Philadelphia, PA

[3]Department of Radiation Oncology, Georgia Regents University, Augusta, GA

[4]Department of Radiation Oncology, Washington University, St. Louis, MO

[5]Department of Radiation Oncology, M.S. Hershey Medical Center, Hershey, PA

[6]Department of Radiation Oncology, Stanford University, Stanford, CA

[7]Department of Radiation Oncology, University of Michigan, Ann Arbor, MI

[8]Department of Radiation Oncology, Case Western University, Cleveland, OH

[9]Department of Radiation Oncology, Princess Margaret Hospital, Toronto, ON, Canada

[10]Radiation Therapy Oncology Group, American College of Radiology, Philadelphia, PA

Abstract

Purpose—To quantify variations in target and normal structure contouring and evaluate dosimetric impact of these variations in non-small-cell lung cancer (NSCLC) cases. To study whether providing an atlas can reduce potential variation.

Methods and Materials—Three NSCLC cases were distributed sequentially to multiple institutions for contouring and radiotherapy planning. No segmentation atlas was provided for the

first two cases (Case1 and Case2). Contours were collected from submitted plans and consensus contour sets were generated. The volume variation among institution contours and the deviation of them from consensus contours were analyzed. The dose-volume histograms (DVH) for individual institution plans were re-calculated using consensus contours to quantify the dosimetric changes. An atlas containing targets and critical structures was constructed and was made available when the third case (Case3) was distributed for planning. The contouring variability in the submitted plans of Case3 was compared with that in first two cases.

Results—Planning Target Volume (PTV) showed large variation among institutions. The PTV coverage in institutions' plans decreased dramatically when re-evaluated using the consensus PTV contour. The PTV contouring consistency did not show improvement with atlas use in Case3. For normal structures, lung contours presented very good agreement, while the brachial plexus showed the largest variation. The consistency of esophagus and heart contouring improved significantly (t-test, $p < 0.05$) in Case3. Major factors contributing to the contouring variation were identified through a survey questionnaire.

Conclusions—The amount of contouring variations in NSCLC cases was presented. Its impact on dosimetric parameters can be significant. The segmentation atlas improved the contour agreement for esophagus and heart, but not for the PTV in this study. Quality assurance of contouring is essential for a successful multi-institutional clinical trial.

Introduction

Lung cancer is the most common cause of cancer death in men and women in the United States [1] with the majority of cases consisting of non-small-cell lung cancer (NSCLC). Radiation therapy plays an important role in the treatment of NSCLC. Three-dimensional (3D) images are utilized to accurately delineate patient anatomy. Highly conformal treatment techniques such as intensity-modulated radiation therapy (IMRT) and volumetric-modulated arc therapy (VMAT) are utilized to deliver the prescribed dose to the planning target volume (PTV) while minimizing dose to organs at risk (OAR). Dose-volume histograms (DVH) are often used in treatment planning to evaluate the plan quality. The dose distribution of conformal radiotherapy plans is very dependent on the contours delineated on patient anatomy.

Studies have shown variation in the contouring of targets and OARs in lung [2, 3, 4, 5, 6], head and neck [7, 8], brain [9, 10], prostate [11, 12], and breast [13, 14] cases. The dosimetric impact of this contour variation can be significant [7, 13, 15] depending on the degree of variation and the plan dose gradient. Differences in structure delineation impact DVH calculation, tumor control probability (TCP), and normal tissue complication probability (NTCP) [11, 12]. Contouring consistency is important in multi-institutional clinical trials to avoid biased trial outcomes derived from inconsistent dosimetric parameters. Various efforts, including atlas construction, protocols for structure definition, and quality assurance (QA) programs, have been developed to reduce contouring inconsistencies and inter-observer variation of target and normal tissue definition in multi-institutional clinical trials [3, 13].

In this study, quantitative analysis of the variation in target and normal structure contouring in NSCLC cases was performed on samples from pre-clinical trial planning studies of RTOG 1106. The dosimetric impact of the contouring variations and the potential of improvement in contouring consistency with a provided atlas were evaluated in conformal radiotherapy plans (including 3D, IMRT, VMAT). Even though a significant amount of literature has reported on the variation of the contouring and the dosimetric impact as referenced above, this report describes this quantity in a multi-institutional clinical trial setting which potentially can be correlated to outcome. Moving the investigations further, we have also collected data on the significant factors that affect the contouring, and have reported on the effect of one of the strategies to reduce contouring variation–the use of an atlas. The need for further improvements is discussed.

Methods and Materials

Data

Three different NSCLC cases were distributed sequentially (Case1, Case2, and Case3) to participating institutions for a pre-clinical trial dry run planning study for Radiation Therapy Oncology Group (RTOG) protocol 1106 [16, 17]. Case1 was a T4N2M0 of the right lower lobe and Case2 was a T4N2M0 of the right upper lobe. These cases were used to study the contouring variation and its impact on dosimetric outcomes. Case3 was a T3N3M0 of the right upper lobe, which was distributed after the Case1 and Case2 studies were completed. Case3 was used to study the impact of providing atlases on contouring consistency. Twelve institutions participated in Case1 dry run, 11 institutions in Case2, and 12 institutions in Case3, with some of the institutions participated in more than one dry run. For each case, the same CT and PET scans were distributed to the participating institutions for contouring and planning. Contouring was completed on the CT image and the PET image was utilized for functional imaging information. Each institution was instructed to contour the gross tumor volume (GTV) to include both primary and nodal tumors present in the CT and PET images. The planning target volume (PTV) was derived by a 1cm expansion of the GTV. OARs included the lungs, heart, esophagus, cord, and brachial plexus. Dose constraints for radiotherapy planning were provided to the institutions with a prescription dose of 74Gy to the PTV and mean lung dose less than 20Gy. 3D conformal, IMRT or VMAT were allowed as planning techniques. All institutions submitted the dry run planning results, but only the datasets with complete DICOM RT files (RT plan, structure, and dose) available at the time of analysis (this included eleven plans for Case1 and seven plans for Case2) were qualified for this study. The volumetric and dosimetric parameters for the same plan calculated by different software systems may differ [18]. Therefore, all submitted plans were imported into a single software system, MIMvista (Version 5.2.1; MIMvista Corp., Cleveland, OH), for consistent analysis.

Contouring variation analysis

Based on the submitted contours for Case1 and Case2, a consensus structure set was generated using the Simultaneous Truth and Performance Level Estimation (STAPLE) algorithm [19] for each individual case and was subsequently reviewed by physicians from participating institutions for further edits and agreement. The contours were also reviewed

and approved by members of Advanced Technology Integration Committee (ATIC) of RTOG to get the final consensus contours. The volume variations between different institutions' contours and the deviation of an individual institution's contour from the consensus contour were analyzed. The deviation of the institution's contour from the consensus contour was evaluated in terms of volume difference, mean surface distance (MSD), and Dice's coefficient, which are defined as the following:

$$\text{Volume difference} = \frac{|X| - |Y|}{|Y|} \times 100\% \quad (1)$$

$$\text{Mean surface distance (MSD)} = \frac{\sum\limits_{x \in \overline{X}} \inf\limits_{y \in \overline{Y}} d(x, y) + \sum\limits_{y \in \overline{Y}} \inf\limits_{x \in \overline{X}} d(x, y)}{\left|\overline{X}\right| + \left|\overline{Y}\right|} \quad (2)$$

$$\text{Dice's coefficient} = \frac{2\,|X \cap Y|}{|X| + |Y|} \times 100\% \quad (3)$$

where X is the set of voxels encompassed by the institution's contour, Y is the set of voxels encompassed by the consensus contour, $|X|$ equals to the number of voxels in set X, $|Y|$ equals to the number of voxels in set Y, $\overline{X}$ is the set of voxels on the surface of contour X, $\overline{Y}$ is the set of voxels on the surface of contour Y, $d(x,y)$ is the Euclidian distance between voxels x and y, inf represents the infimum, and symbol $\cap$ represents the intersection between two sets.

Dosimetric impact of contouring variation

Dose and structures submitted by each institution were imported into MIMvista and DVHs were calculated. The institution contours were replaced by the consensus contours and the DVHs were re-calculated for each plan to quantify the dosimetric changes caused by contouring deviations. Representative dose points were evaluated, which included PTV D95%, PTV V100%, cord maximum dose, mean lung dose, lung V20Gy, mean esophagus dose, heart V65Gy, and brachial plexus maximum dose. The TCP for each plan was calculated using a logistic model [20]

$$\text{TCP} = \frac{1}{1 + \left(\frac{D_{50}}{EUD}\right)^{4\gamma}} \quad (4)$$

EUD where D_{50} denotes the PTV dose required for a 50% probability of tumor control and γ denotes the normalized slope of the sigmoid-shaped dose-response curve at D_{50}. Equivalent uniform dose (EUD) is calculated using following equation

$$EUD = \left(\frac{1}{N} \sum_i D_i^a\right)^{\frac{1}{a}} \quad (5)$$

where N is the number of voxels in the anatomic structure of interest, D_i is the dose in the i th voxel, and a is the tumor or normal tissue-specific parameter that describes the dose-volume effect. The parameters, D_{50}, γ, and a, vary with tumor type, treatment method, etc., and a set of numbers from relevant literatures [21, 22] were used to calculate TCP in this study. A comparison was performed between the TCP values calculated from the institution's PTV DVH and the consensus PTV DVH.

Impact of an atlas on contouring consistency

Atlases containing targets and OARs were constructed based on the consensus contours from Case1 and Case2. The atlases were made available on the RTOG website (http://www.rtog.org/CoreLab/ContouringAtlases/LungAtlas.aspx) when Case3 was distributed for planning. Participating institutions were instructed to refer to the atlases when contouring, and the remaining components of radiotherapy planning remained similar as in Case1 and Case2. All twelve plans from participating institutions were collected for Case3 and the contouring variation was quantified using the same analysis described in Equations (1)–(3). The calculated numbers were then compared with those derived from Case1 and Case2.

Major factors contributing to contouring variation

To identify possible causes of the observed contouring variations, a questionnaire was made and sent to participating physicians who contoured the dry run cases to find out what they consider to be major factors that contributed to the contouring variability. The factors included in the questionnaire are listed in Table 1. The physicians were asked to score these factors as to how much they impact the contouring. Scores were given from 0 to 5 with score 0 as having no impact and score 5 as having the most significant impact.

Results

Contouring variations in Case1 and Case2

Figure 1 shows the GTV contour variation among multiple institutions for a single axial CT slice with the consensus contour displayed in red.

Quantification of the variation is shown in Table 2 and Table 3. The volume difference between the institution's PTV and consensus PTV ranged from −28.0% to 7.7%, with a mean value of −5.9 and standard deviation (SD) of 10.75% in Case1. The PTV volume difference was more pronounced in Case2 ranging from −40.8% to 19.9% and mean±SD of −10.3%±22.3%. The negative mean values of PTV volume difference suggest that the institutions tended to under-contour the treatment target in both cases (Fig. 1). For normal structures, brachial plexus presented a large variation in contouring (Dice's coefficient below 50%), esophagus and heart presented less variation (Dice's coefficient around 75% for esophagus and around 85% for heart), and lungs presented the least amount of variation (Dice's coefficient above 95%). The numbers for cord in Table 3 do not represent the real variation in contouring cord because the cord was contoured differently in length by different institutions.

Dosimetric impact of contouring variation

The dose-volume parameters calculated with the institutions' contours and the consensus contours are listed in Table 2 and Table 3. PTV coverage decreased when evaluated using the consensus contour in both cases, especially with dramatic changes in the DVH in Case2 where the PTV contouring variation was more pronounced. The mean value of TCP decreased from 84.3%/87.8% (Case1/Case2) to 72.6%/52.8% when consensus PTV was used for re-calculation. Despite the brachial plexus exhibiting one of the largest contouring variations, the dose to the brachial plexus was minimal in all plans for both cases and the variation did not have a significant impact on plan quality. The maximum cord dose increased in both cases when the consensus cord contours were used for dose re-calculation.

Figure 2 shows the PTV DVHs for Case2 calculated with the institutions' contours and the consensus contour respectively. Some of the DVH curves showed severe degradation depending on the deviation of PTV volume from the consensus contour. Figure 3 shows the correlation between PTV volume differences and dosimetric outcome changes. As expected, larger decreases in PTV volume compared to the consensus PTV resulted in greater degradation of standard PTV DVH metrics (D95%, etc.). If the PTV was over-contoured compared to the consensus PTV, it certainly did not affect the dosimetric outcomes in terms of PTV coverage. However, with over-contoured PTVs, the treatment plan has the potential to treat a greater volume of normal tissue to prescription dose than it would do with the consensus contour. The mean value of lung V20Gy in over-contoured-PTV plans was 36.5% ±3.1% (mean±SD) compared to 31.9%±4.0% of under-contoured-PTV plans in Case1, and the values were 33.4%±0.5% vs. 27.8%±3.9% in Case2. The dependency of dosimetric outcomes with respect to Dice's coefficient and MSD was similar to that with respect to PTV volume difference.

Impact of an atlas on contouring consistency

Table 4 quantifies contouring variations in Case3 (use of an atlas) in comparison to Case1 and Case2. PTV volumes were 456cc±52cc (mean±SD), 514cc±127cc, and 568cc±67cc in Case1, Case2, and Case3, respectively. The Dice's coefficient of PTV in Case3 was 88.3% ±2.2%, which did not show improvement from Case1 (92.4%±3.3%) and Case2 (86.4% ±7.6%). The Dice's coefficients of esophagus and heart in Case3 significantly improved (t-test, $p<0.05$) from those in Case1 and Case2. The brachial plexus did not show significant improvement in Dice's coefficient but the mean surface distance reduced dramatically in Case3. Dice and mean surface distance evaluations for cord were not performed due to different lengths contoured from different institutions.

Major factors contributing to contouring variation

Among the factors included in the questionnaire for tumor contouring, four had average scores larger than 3 (total score 5), which were: difficulty in defining tumor vs. other pathologic structures (e.g. atelectasis, infection, condensation, effusion) (average score 3.67); difficulty in defining tumor vs. normal structures (e.g. mediastinal vessels) (average score 3.33); involvement of lymph nodes (average score 3.17); and variation in SUV threshold for tumor delineation in PET image (average score 3.17). For normal tissue contouring, CT image quality (e.g. resolution, contrast, partial volume effect) (average score

3.67) and physician's knowledge and judgment (average score 3.33) werethe two leading factors.

Discussion

This study examined the magnitude of variations in target and normal tissue contouring in NSCLC radiotherapy. The results showed quite large variability among radiation oncologists in contouring target and some critical structures. A similar amount of variation for NSCLC cases was also reported in previous studies [3, 4]. While only volume and dimensional differences were evaluated in these studies, our work quantified more parameters including Dice's coefficient and mean surface distance, and correlated these variations with changes in radiation dose to target and OARs. The dosimetric impact of contour variation can be significant. The dosimetric degradation in this study was especially pronounced for PTVs which exhibit a high dose gradient on the boundary compared to other structures in highly conformal radiotherapy plans. Due to steep falloff of dose outside the PTV, portions of the treatment target that are excluded from the PTV contour may receive significantly lower radiation dose. As can be seen from Figure 2, the low dose part of the PTV DVH curve is significantly affected by the under-contoured PTV which in turn decreases the equivalent uniform dose and TCP. The dosimetric impact of OAR contouring variation is dependent on the proximity of the OAR to the target and the dose gradient in the OAR region. In general, OAR dosimetry is less affected by contouring in the observed variation range. However, accurate dose statistics are critical for clinical trial outcomes analysis (e.g. NTCP modeling), although variation within a certain range may not have a significant clinical impact on individual cases.

Another aspect that distinguishes this study from others is the investigation of the role of atlases in contouring target and OARs. The third NSCLC case (Case3), which had a similar complexity to the first two cases, showed comparable variation in target contouring despite the availability of segmentation atlases. The OAR contours which displayed either severe or moderate variations in first two cases presented considerable improvement in consistency with the use of atlases in Case3. A probable cause of the PTV inconsistency is that the target has a large variation in size, location, pathology, and nodal involvement among different cases while the normal organs or tissues do not have such large variations.

This study was limited by the small number of cases. Since this was part of the pre-clinical trial planning study, each dry run took long periods from the selection of a case to collection of plans. The third case (Case3) was the last dry run of the RTOG 1106 pre-trial planning study and was used as the credentialing benchmark in this trial. Although the number of cases was small, the treatment plans included in this study were collected from a large number of institutions, which made these datasets particularly valuable for inter-institutional comparison study.

The possible causes of contouring variation were discussed in the study by Van de Steene *et al* [4]. Studies from Steenbakkers *et al* [5] and Fitton *et al* [6] found that using PET in addition to CT reduces interobserver variability in delineation of lung target volumes, especially for the targets in hilar region, heart, great vessels, pericardium, mediastinum,

and/or the region associated with atelectasis [6]. Nevertheless, the remaining variation was still large compared with other geometric uncertainties such as setup variation and organ motion [5]. Based on existing studies and experience from current work, we developed a survey questionnaire to collect contouring physicians' opinions on contributing factors to the contouring variability. Among a number of confounding factors in tumor delineation variability, the discrepancy in defining tumor versus atelectasis and mediastinal vessels seems to be a major factor in our study. Even though PET images were provided, the presence of nodal volumes added variability in GTV contouring. Identifying the causes of contour variations will help us to come up with more targeted solutions and to improve contouring consistency in future trials. For example, contouring uncertainties caused by inferior image quality can be reduced through credentialing of institution's imaging devices, which is already under consideration in some of the RTOG protocols.

The contour deviation demonstrated in this study indicated the possibility for poor dosimetric compliance to the protocol in clinical trials thus having an adverse impact on clinical trial outcomes. A properly formed CTV to PTV margin which incorporates the delineation uncertainty may eliminate the high dose gradients from true CTV boundary during treatment [23]. However, this may not always be feasible since allowing PTV margins to be too large may have negative impact in the clinical trial setting. The greatest impact on dosimetry in this study was from variation in tumor contouring which did not appear to benefit from the provided atlases. PTV volume also affected dose distribution in surrounding tissues significantly. Efforts beyond atlas construction should be implemented in clinical trials to improve contouring consistency, which may include detailed and specific contouring instructions as part of clinical trial protocol, educational workshops, and remote case reviews. Providing atlas is one method that was evaluated in this work and certainly more studies need to be done in this area. Quality assurance of contours for the cases included in the trial is essential, and if necessary, pre-trial dry runs and quality reviews are also desirable to identify issues and assess the capability of a successful trial implementation.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA: A Cancer Journal for Clinicians. 2012; 62:10–29. [PubMed: 22237781]

2. Giraud P, Elles S, Helfre S, et al. Conformal radiotherapy for lung cancer: different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists. Radiotherapy and Oncology. 2002; 62:27–36. [PubMed: 11830310]

3. Spoelstra FOB, Senan S, Le Péchoux C, et al. Variations in Target Volume Definition for Postoperative Radiotherapy in Stage III Non-Small-Cell Lung Cancer: Analysis of an International Contouring Study. Int J Radiat Oncol Biol Phys. 2010; 76:1106–1113. [PubMed: 19560881]

4. Van de Steene J, Linthout N, de Mey J, et al. Definition of gross tumor volume in lung cancer: inter-observer variability. Radiotherapy and Oncology. 2002; 62:37–49. [PubMed: 11830311]

5. Steenbakkers RJ, Duppen JC, Fitton I, et al. Reduction of observer variation using matched CT-PET for lung cancer delineation: A three-dimensional analysis. Int J Radiat Oncol Biol Phys. 2006; 64:435–448. [PubMed: 16198064]

6. Fitton I, Steenbakkers RJ, Gilhuijs K, et al. Impact of anatomical location on value of CT-PET co-registration for delineation of lung tumors. Int J Radiat Oncol Biol Phys. 2008; 70:1403–1407. [PubMed: 17980511]

7. Nelms BE, Tomé WA, Robinson G, et al. Variations in the Contouring of Organs at Risk: Test Case From a Patient With Oropharyngeal Cancer. Int J Radiat Oncol Biol Phys. 2012; 82:368–378. [PubMed: 21123004]

8. Hall WH, Guiou M, Lee NY, et al. Development and Validation of a Standardized Method for Contouring the Brachial Plexus: Preliminary Dosimetric Analysis Among Patients Treated With IMRT for Head-and-Neck Cancer. Int J Radiat Oncol Biol Phys. 2008; 72:1362–1367. [PubMed: 18448267]

9. Cattaneo GM, Reni M, Rizzo G, et al. Target delineation in post-operative radiotherapy of brain gliomas: Interobserver variability and impact of image registration of MR(pre-operative) images on treatment planning CT scans. Radiotherapy and Oncology. 2005; 75:217–223. [PubMed: 15885831]

10. Yamamoto M, Nagata Y, Okajima K, et al. Differences in target outline delineation from CT scans of brain tumours using different methods and different observers. Radiotherapy and Oncology. 1999; 50:151–156. [PubMed: 10368038]

11. Rosewall T, Bayley AJ, Chung P, et al. The effect of delineation method and observer variability on bladder dose-volume histograms for prostate intensity modulated radiotherapy. Radiotherapy and Oncology. 2011; 101:479–485. [PubMed: 21864921]

12. Foppiano F, Fiorino C, Frezza G, et al. The impact of contouring uncertainty on rectal 3D dose-volume data: Results of a dummy run in a multicenter trial (AIROPROS01-02). Int J Radiat Oncol Biol Phys. 2003; 57:573–579. [PubMed: 12957271]

13. Li XA, Tai A, Arthur DW, et al. Variability of Target and Normal Structure Delineation for Breast Cancer Radiotherapy: An RTOG Multi-Institutional and Multiobserver Study. Int J Radiat Oncol Biol Phys. 2009; 73:944–951. [PubMed: 19215827]

14. Lorenzen EL, Taylor CW, Maraldo M, et al. Inter-observer variation in delineation of the heart and left anterior descending coronary artery in radiotherapy for breast cancer: A multi-centre study from Denmark and the UK. Radiotherapy and Oncology. 2013; 108:254–258. [PubMed: 23891098]

15. Hernandez SY, Magliari J, Lin J, et al. Normal Tissue Contour Variation and the Dosimetric Impact on Radiation Therapy Treatment Planning. Int J Radiat Oncol Biol Phys. 2011; 81:S804.

16. Kong FM, Machtay M, Videtic G, et al. Using during RT PET to Individualize Adaptive RT for Patients with Stage III NSCLC: A RTOG Planning Study [Abstract]. Int J Radiat Oncol Biol Phys. 2011; 81:S603.

17. Xiao Y, Cui Y, Kong FM, et al. Initial Experience With A Pre-Clinical Trial Process for A Dosimetric Feasibility Study of Adaptive Radiotherapy for Patients With Stage III NSCLC[Abstract]. Int J Radiat Oncol Biol Phys. 2011; 81:S613.

18. Ackerly T, Andrews J, Ball D, et al. Discrepancies in volume calculations between different radiotherapy treatment planning systems. Australasian Physical & Engineering Science in Medicine. 2003; 26:90–92.

19. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging. 2004; 23:903–921. [PubMed: 15250643]

20. Martel MK, Ten Haken RK, Hazuka MB, et al. Estimation of tumor control probability model parameters from 3-D dose distributions of non-small cell lung cancer patients. Lung Cancer. 1999; 24:31–37. [PubMed: 10403692]

21. Mehta N, King CR, Agazaryan N, et al. Stereotactic body radiation therapy and 3-dimensional conformal radiotherapy for stage I non-small cell lung cancer: A pooled analysis of biological equivalent dose and local control. Practical Radiation Oncology. 2012; 2:288–295. [PubMed: 24674167]

22. Zhao L, Chen M, Ten Haken R, et al. Three-dimensional conformal radiation may deliver considerable dose of incidental nodal irradiation in patients with early stage node-negative non-small cell lung cancer when the tumor is large and centrally located. Radiotherapy and Oncology. 2007; 82:153–159. [PubMed: 17287040]

23. van Herk M. Errors and margins in radiotherapy. Semin Radiat Oncol. 2004; 14:52–64. [PubMed: 14752733]
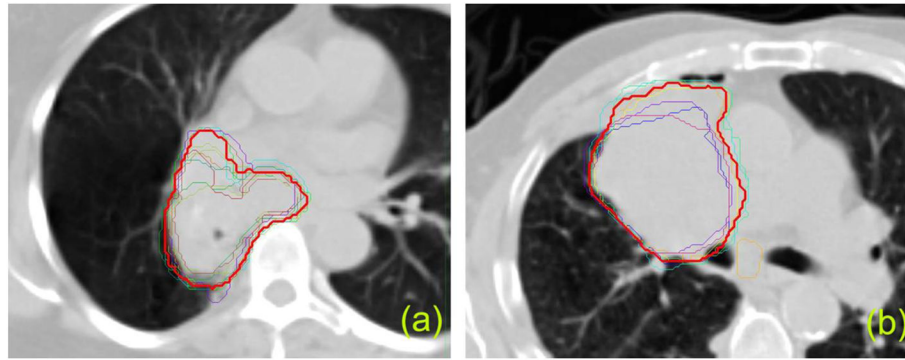
Fig. 1.
GTV contours from different institutions. Red thick line represents the consensus contour.
(a) Case1, (b) Case2.
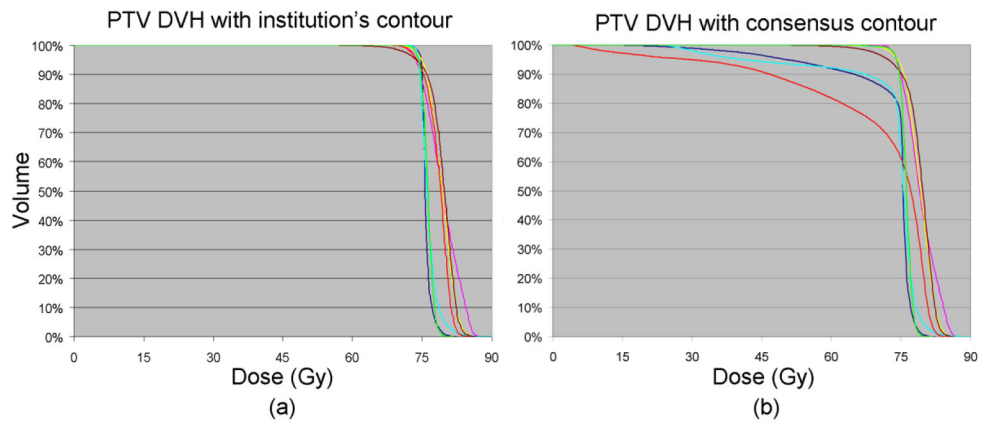
Fig. 2.
DVH curves of PTV for Case2. (a) DVH calculated with institutions' contours and institutions' planned doses. (b) DVH calculated with consensus contour and institutions' planned doses.

Fig. 3.

Dosimetric outcome changes due to PTV contouring variation. X axis represents the deviation of institutions' PTV volume from consensus PTV volume, and Y axis represents the deviation of PTV dose parameters calculated with consensus contour from the parameters calculated with institutions' contours. (a) Case1, (b) Case2.

Table 1

Questionnaire - Factors contributed to contouring variation in RTOG 1106 dry run cases
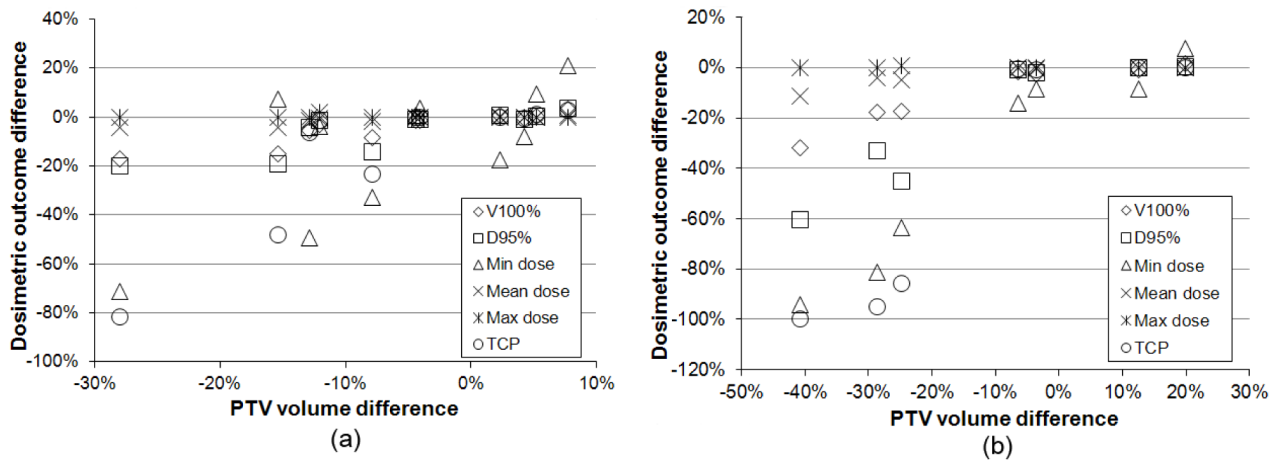
For tumor (including both primary and nodal tumors) contouring:
CT image quality (e.g. resolution, contrast, partial volume effect)
PET image quality (e.g. resolution, contrast, partial volume effect)
Uncertainties associated with the contouring tools/devices (drawing precision)
Imperfect window setting (i.e. window width and level) when contouring
Variation in SUV threshold for tumor delineation in PET image
Time spent on contouring (diligence in contour drawing)
Difficulty in defining tumor vs. some other pathologic structures (e.g. atelectasis, infection, condensation, effusion)
Difficulty in defining tumor vs. some normal structures (e.g. mediastinal vessels)
Involvement of lymph nodes
Unclear or not-so-detailed instructions and definitions in the protocol
Physician's knowledge and judgment
Other factors (specify)
For normal tissue/OAR contouring:
CT image quality (e.g. resolution, contrast, partial volume effect)
Uncertainties associated with the contouring tools/devices (drawing precision)
Imperfect window setting (i.e. window width and level) when contouring
Time spent on contouring (diligence in contour drawing)
Unclear or not-so-detailed instructions and definitions in the protocol
Physician's knowledge and judgment
Other factors (specify)

Table 2

Summary of volume and dose parameters for PTV

Structure	Quantity	Case1 (*n* = 11)				Case2 (*n* = 7)			
		min	max	mean	SD	min	max	mean	SD
PTV	Volume (cc)	349.0	522.4	456.1	52.1	339.1	686.4	513.9	127.4
	MSD* (mm)	1.91	4.40	2.55	0.68	2.31	8.07	4.56	2.18
	Dice[†]	83.6%	95.4%	92.4%	3.3%	74.7%	94.5%	86.4%	7.6%
	D95% (Gy) - inst[‡]	59.01	72.97	69.86	3.99	73.69	74.89	74.20	0.49
	D95% (Gy) - cons[§]	47.78	72.42	66.34	7.74	29.36	74.25	59.27	18.93
	V100% - inst	69.9%	93.1%	86.9%	6.9%	93.7%	98.3%	96.0%	1.7%
	V100% - cons	59.3%	92.2%	83.2%	9.4%	64.5%	95.6%	86.4%	11.7%
	TCP - inst	66.0%	88.8%	84.3%	6.5%	85.5%	89.5%	87.8%	1.9%
	TCP - cons	16.1%	89.0%	72.6%	24.7%	0.00%	89.4%	52.8%	44.4%

Abbreviations: PTV = planning target volume; MSD = mean surface distance; Dice = Dice's coefficient; TCP = tumor control probability; *n* = number of cases.

*
Mean surface distance between institutions' contour and consensus contour.

[†]
Dice's coefficient between institutions' contour and consensus contour.

[‡]
Dose parameter calculated with institutions' contour.

[§]
Dose parameter calculated with consensus contour.

Table 3

Summary of volume and dose parameters for normal tissues

Structure	Quantity	Case1 (*n* = 11)		Case2 (*n* = 7)	
		mean	SD	mean	SD
Cord	Volume (cc)	24.76	5.48	54.59	18.63
	Dice *	81.3%	6.9%	76.2%	8.6%
	Max dose (Gy) - inst [†]	47.79	3.19	35.31	9.69
	Max dose (Gy)–cons [‡]	51.08	4.33	36.60	10.76
Lungs	Volume (cc)	2258	49	3912	242
	Dice	97.5%	1.0%	96.9%	1.6%
	V20Gy - inst	33.8%	3.9%	29.8%	4.0%
	V20Gy - cons	33.6%	4.2%	29.4%	4.2%
Esophagus	Volume (cc)	25.41	6.94	46.06	14.53
	Dice	77.3%	9.3%	75.8%	11.0%
	Mean dose (Gy) - inst	39.26	6.69	18.35	4.69
	Mean dose (Gy) - cons	39.91	3.66	19.45	4.98
Heart	Volume (cc)	490.4	77.7	896.6	176.7
	Dice	86.4%	6.1%	86.8%	5.1%
	V65Gy - inst	18.9%	8.7%	2.47%	3.42%
	V65Gy - cons	20.7%	9.6%	2.43%	2.53%
Brachial Plexus	Volume (cc)	13.76	4.90	16.01	7.09
	Dice	40.5%	17.4%	40.4%	21.2%
	Max dose (Gy) - inst	1.78	0.71	2.25	0.76
	Max dose (Gy) - cons	1.92	0.65	1.98	0.47

Abbreviations: Dice = Dice's coefficient; *n* = number of cases.

*
Dice's coefficient between institutions' contour and consensus contour.

[†]
Dose parameter calculated with institutions' contour.

[‡]
Dose parameter calculated with consensus contour.

Table 4

Comparison of contouring consistency in Case3 with that in Case1 and Case2

| Quantity | PTV | | Esophagus | | Heart | | Brachial Plexus | |
Case	MSD* (mm)	Dice†	MSD (mm)	Dice	MSD (mm)	Dice	MSD (mm)	Dice
Case1 (n = 11)	2.55	92.4%	2.16	77.3%	4.45	86.4%	16.27	40.5%
Case2 (n = 7)	4.56	86.4%	2.65	75.8%	3.85	86.8%	19.85	40.4%
Case3 (n = 12)	3.02	88.3%	1.62	84.2%	2.38	93.0%	8.09	42.97%

Abbreviations: PTV = planning target volume; MSD = mean surface distance; Dice = Dice's coefficient; *n* = number of cases.

*
Mean surface distance between institutions' contour and consensus contour.

†
Dice's coefficient between institutions' contour and consensus contour.