Interobserver delineation variation in lung tumour stereotactic body radiotherapy

¹G F PERSSON, MD, PhD, ¹D E NYGAARD, MSc, ^{1,2,3}C HOLLENSSEN, MSc, ¹P MUNCK af ROSENSCHÖLD, PhD, ⁴L S MOURITSEN, MD, ¹A K DUE, MD, ^{1,2}A K BERTHELSEN, MD, ^{1,5}J NYMAN, MD, PhD, ^{1,2}E MARKOVA, MD, ¹A P ROED, BSc, ¹H ROED, MD, DMSc, ^{1,6,7}S KORREMAN, PhD and ^{1,8}L SPECHT, MD, DMSc

¹Department of Radiation Oncology, Copenhagen University Hospital—Rigshospitalet, Copenhagen, Denmark,

²Department of Clinical Physiology and Nuclear Medicine and PET, Copenhagen University Hospital—Rigshospitalet, Copenhagen, Denmark, ³Department of Informatics and Mathematical Models, Technical University of Denmark, Copenhagen, Denmark, ⁴Department of Radiation Oncology, Copenhagen University Hospital—Herlev Hospital, Denmark,

⁵Department of Oncology, Sahlgrenska University Hospital, Gothenburg, Sweden, ⁶Department of Science, Systems and Models, Roskilde University, Denmark, ⁷Faculty of Sciences, Niels Bohr Institute, University of Copenhagen, Denmark, ⁸Faculty of Health Sciences, Copenhagen University, Denmark

Objectives: In radiotherapy, delineation uncertainties are important as they contribute to systematic errors and can lead to geographical miss of the target. For margin computation, standard deviations (SDs) of all uncertainties must be included as SDs. The aim of this study was to quantify the interobserver delineation variation for stereotactic body radiotherapy (SBRT) of peripheral lung tumours using a cross-sectional study design.

Methods: 22 consecutive patients with 26 tumours were included. Positron emission tomography/CT scans were acquired for planning of SBRT. Three oncologists and three radiologists independently delineated the gross tumour volume. The interobserver variation was calculated as a mean of multiple SDs of distances to a reference contour, and calculated for the transversal plane (SD_{trans}) and craniocaudal (CC) direction (SD_{cc}) separately. Concordance indexes and volume deviations were also calculated.

Results: Median tumour volume was 13.0 cm³, ranging from 0.3 to 60.4 cm³. The mean SD_{trans} was 0.15 cm (SD 0.08 cm) and the overall mean SD_{cc} was 0.26 cm (SD 0.15 cm). Tumours with pleural contact had a significantly larger SD_{trans} than tumours surrounded by lung tissue.

Conclusions: The interobserver delineation variation was very small in this systematic cross-sectional analysis, although significantly larger in the CC direction than in the transversal plane, stressing that anisotropic margins should be applied. This study is the first to make a systematic cross-sectional analysis of delineation variation for peripheral lung tumours referred for SBRT, establishing the evidence that interobserver variation is very small for these tumours.

Received 1 June 2011
Revised 22 August 2011
Accepted 12 September 2011

DOI: 10.1259/bjr/76424694

© 2012 The British Institute of Radiology

Target delineation is overall a major source of uncertainty in radiotherapy as it contributes to systematic errors and can lead to geographical miss of the target. The true gross tumour volume (GTV) in the individual radiotherapy patient is unknown; hence the delineation uncertainty can only be estimated. An often-used approach is to examine the delineation variation in a group of observers delineating the same target [1–5].

Delineation variation depends on the target, the surrounding tissue, the imaging modality, the observers and the delineation protocol. In a study by Steenbakkers

et al [2], analysing delineation variation for locally advanced lung cancer with and without ¹⁸F-fluorodeoxyglucose positron emission tomography (FDG-PET) information, it was found that the addition of FDG-PET decreased delineation variation and that delineation variation was much larger where the tumour adjoined the mediastinum or atelectasis than where the tumour adjoined the lung or the chest wall. The integration of FDG-PET/CT into radiotherapy planning has overall had a major impact on delineation variation. The addition of metabolic information alters radiotherapy planning in two different ways: by the detection of additional lymph nodes or distant metastasis, thereby altering the disease staging; and by helping to distinguish between tumour and surrounding tissue with similar density in the GTV delineation process (e.g. the mediastinum or atelectasis), altering the size of the GTV [2, 6–10]. This undoubtedly produces a positive impact, with a decrease in delineation variation in

Address correspondence to: Dr Gitte Fredberg Persson, Department of Radiation Oncology, Section 3994, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark. E-mail: gitte.persson@rh.regionh.dk

This work has been supported by grants from the following institutions: the Danish Council for Independent Research, the Lundbeck Foundation Center for Interventional Research in Radiation Oncology (CIRRO) and the Arvid Nilsson Foundation.

advanced tumours with mediastinal involvement and atelectasis.

Results from multicentre studies may not be relevant for single centres as, optimally, interobserver delineation variation has to be determined within each centre, for all planning protocols and all patient groups.

Clinical margins can be calculated using a probabilistic margin formula, as presented by van Herk et al [11]. The margin formula takes into account systematic uncertainties, random uncertainties and the width of the beam penumbra. All uncertainties are added as standard deviations (SDs) and the systematic uncertainties are weighted higher than the random; that is, in order to calculate an appropriate margin for creation of the planning target volume (PTV), all uncertainties have to be quantified as SD [11–13].

The primary aim of this cross-sectional study was to evaluate the interobserver delineation variation for peripheral lung tumours referred for stereotactic body radiotherapy (SBRT) and planned using dedicated FDG-PET/CT and a well-defined delineation protocol. Because of the anisotropic image resolution of CT scans for radiotherapy planning, we hypothesise that the interobserver delineation variation is larger in the craniocaudal (CC) direction than in the transversal plane.

In addition, we compared (group-wise) the variation in GTV delineations performed by oncologists and radiologists, and we compared the interobserver delineation variation for tumours with and without pleural contact.

Methods and materials

Data collection for this project was approved by the Data Protection Agency (ref. no. 2007-58-0015). All consecutive patients receiving SBRT of lung tumours in the Department of Radiation Oncology at Copenhagen University Hospital Rigshospitalet in 2008 were eligible for the study. Patients were referred from clinics in eastern Denmark. According to the clinical protocol in the department, only patients with T1–2 lung tumours who were inoperable owing to medical conditions or patients with lung metastases were eligible for SBRT, and they could have no more than three tumours, with a maximum diameter of 6 cm. The clinical protocol at that time prescribed a FDG-PET/CT with intravenous contrast enhancement for radiotherapy planning. The patients rested for 30 min before injection of 400 MBq FDG. After the injection the patients rested for a further 60 min before the FDG-PET/CT scans were acquired on a Siemens Biograph™ FDG-PET/CT scanner (Siemens Medical Solutions, Erlangen, Germany). The scanner has 24 detectors, but only the central 16 are used for the FDG-PET/CT scan. Each detector measures 0.75 mm, giving a collimation of 12 mm in total. CT scans were performed in a helical mode with a pitch of 1.2. The scans were reconstructed with a slice thickness of 3 mm. In one case a patient had only a conventional CT performed, and in this case the CT scanner was a Siemens Sensation Open scanner (Siemens Medical Solutions) with a collimation of 24 detectors, each measuring 1.2 mm. A helical scan mode with a pitch of 1.2 was used. The scan was reconstructed with a slice thickness of 2.5 mm. The rotation time was 1 s for both scanners. The patients were immobilised with

arms above their head using a VacFix® vacuum bag (PAR Scientific A/S, Odense, Denmark) in a Styrofoam (Dow Chemical Company, Midland, MI) shell.

Delineations

As a part of clinical routine, the FDG-PET/CT scans were analysed by a nuclear medicine physician and a radiologist as a first step in the planning process. The FDG-PET/CT scan was performed as a whole body scan and was systematically analysed to assure that the disease had not metastasised or developed further. The PET positive lesions were then roughly marked by a contour performed by the nuclear medicine physician. Only the CT scan and this contour (and thus not the PET scan itself) were imported to Eclipse (Varian Medical Systems, Palo Alto, CA), where the GTV delineation was performed. The PET contour was not a reference contour and was used only as a pointer to where the lesion was. The delineation protocol prescribed that the contouring of GTV was performed on the CT scan and only in cases of atelectasis or if the tumour was difficult to distinguish from mediastinal structures the rough PET contour was used to delimitate the GTV. This is rarely the case with peripheral lung tumours.

In our clinical setting the delineation is performed by an oncologist together with a radiologist. For the present study the CT and the PET contours were exported to a separate Eclipse research database. Delineations of GTV were performed independently by six observers, with a fixed broad lung window (–1000 to 700 HU). The magnification factor was left to the observer to decide. The observers were asked to delineate the GTV without changing the window setting and according to ICRU 50 guidelines specifying “the gross visible extent and location of the malignant growth” [14], and thus not including any assumed uncertainties. The observers used the tools they were most comfortable with to delineate: four observers used a computer mouse to delineate and two observers used a Cintiq tablet (Wacom, Otone, Japan) to delineate directly on a pressure-sensitive screen. Three clinical oncologists and three radiologists delineated the 26 tumours in the 22 patient CT scans. All observers had participated in dedicated target delineation workshops and were used to work together in radiologist/oncologist pairs for the target delineation process. The experience in the oncologist group ranged from 4 to 12 years while the experience in the radiologist group ranged from 1 to 18 years. All observers worked in one of the two radiotherapy centres in eastern Denmark. A written protocol was given to all observers, describing the procedure of opening the patient scans in a blinded way, to ensure an independent delineation process. Observers were asked to delineate without changing the window setting. All observers were familiar with the process of reviewing the contours in the coronal and sagittal planes. The observers had no access to the medical records of the patients.

Data analysis

The dicom data of the six contours for each tumour were extracted from Eclipse and analysed in MATLAB

version 2007b (MathWorks™, Natick, MA) software program. The program was tested using a constructed dicom test data set consisting of cylinders with different pre-set diameters.

Several parameters expressing the interobserver delineation variation were computed. Concordance indexes (CI) were calculated as relative measures of interobserver delineation variation, and SDs of distances between contours were calculated as absolute measures of interobserver delineation variation. Different methods were used to analyse the interobserver delineation variation in the transversal plane and the CC direction since the interobserver delineation variation in the CC direction is more dependent on the slice thickness. The relative and absolute interobserver delineation variations were compared for tumours with and without pleural contact, and also among radiologists and oncologists.

Analysis of interobserver delineation variation in the craniocaudal direction

The mean position of the centre of volume (CoV) of the six contoured volumes for each tumour was calculated. From this point the distances to the most cranial slice of each of the six contours were measured and a mean distance calculated. At this mean distance a reference cranial end-plane was defined. The same was done in the caudal direction, where a reference caudal end-plane was defined. The distances from the most cranial slice of each contour to the reference cranial end-plane (six distances) and from the most caudal plane of each contour to the reference caudal end-plane (six distances) were calculated. The SD of these 12 distances (SD_{cc}) was considered a measure of the tumour-specific interobserver delineation variation in the CC direction. The mean of all tumour-specific SD_{cc} was considered the overall interobserver delineation variation in the CC direction. For calculations within the subgroups of radiologists and oncologists only the three contours of radiologists and oncologists, respectively, were used, and new CoVs and reference end-planes were calculated.

Analysis of interobserver delineation variation in the transversal plane

For each slice the CoV of all six contours was calculated and used as a reference point. A mean contour was created for each slice. The co-ordinates for the mean contour was calculated as the mean distances from the reference point to the contours in 360 equally spaced angles. Figure 1 shows a schematic example of a slice with contours, reference point and mean contour. A minimum of two contours had to be present in each slice in order to calculate a mean contour. In the 360 equally spaced angles, distances from the delineated contours to the mean contour were calculated. In case the contours intersected with the angles more than once, a distance for each intersection was measured. For each angle a local SD of the distances to the mean contour was calculated. The mean of all local SDs was considered a measure for the patient-specific interobserver delineation variation in the transversal plane (SD_{trans}). The mean of

all patient-specific SD_{trans} was considered a measure of the overall interobserver delineation variation. For calculations within the subgroups of radiologists and oncologists, only the three contours of radiologists and oncologists, respectively, were used, and new reference points and mean contours were calculated.

The tumour volume was calculated as the mean of the GTVs delineated by the six observers. Correlation analysis was performed for tumour volume and SD_{trans} and SD_{cc} , respectively.

Concordance index

CI is used as a relative measure of interobserver delineation variation and was calculated for each tumour and for each group of observers with two different methods.

The tumour-specific CI_{all} is calculated as the ratio between the intersecting volume and the common volume of all the contours by the different observers:

$$CI_{all} = \frac{A \cap B \cap C \cap D \cap E \cap F}{A \cup B \cup C \cup D \cup E \cup F} \quad (1)$$

where A, B, C, D, E and F represent GTVs of the six observers. The overall CI_{all} was the mean of all tumour-specific CI_{all} . The tumour-specific $CI_{pairwise}$ was calculated as the mean of intersecting/common volume ratios for all

Figure 1. Schematic figure showing the contours delineated by the six observers on a single slice. Each colour represents an observer and the dots represent the centre of volume (CoV) of each contour. A reference point is calculated as the mean position of the CoVs (black dot). A mean contour is calculated (black line) as the mean distance from the reference point to the six contours in 360 equally spaced angles radiating from the reference point. A magnification is inserted in the right lower corner: the distances from each contour to the mean contour, in 360 equally spaced angles radiating from the reference point, were measured and a local standard deviation calculated.

pairwise combinations of observers per tumour (15 values). The overall CI_{pairwise} was the mean of all tumour-specific CI_{pairwise} . For calculations within the subgroups of radiologists and oncologists only the three contours of radiologists and oncologists, respectively, were used.

In both cases a value of 1 would indicate complete agreement among the observers and a value of 0 would indicate complete disagreement among the observers. The CI_{all} value is very sensitive to even small variations in overlap in volume (*i.e.* assuming there are variations in the delineation), and the CI_{all} will tend to decrease with an increasing number of observers. The mean CI_{pairwise} is a more robust parameter for interobserver variation and is not influenced by the number of observers.

As delineation uncertainty concerns the surface of tumours, it can be hypothesised that the relative magnitude of the delineation error will be larger for the small tumours than for the larger tumours. We tested for a correlation between tumour volume and CI_{pairwise} .

Volume analysis

To examine what the impacts of delineation uncertainty were on delineated GTV size, the GTV sizes of all observers for all tumours were measured by a software tool in Eclipse. For each tumour the absolute and relative difference between the largest and the smallest delineated GTV size was calculated.

Statistics

Student's *t*-test for paired and unpaired samples was used for comparisons of data (two-sided level of significance, $p < 0.05$). All correlations were calculated as linear regression analysis with calculation of Pearson correlation coefficient (*r*) and 95% confidence intervals (CI).

The VassarStats website [15] and the SPSS® statistics program (v. 17.0; SPSS, Chicago, IL) were used for statistical computations.

Results

27 patients were eligible for the study. Five patients (three with primary lung cancer and two with lung metastases) were excluded as their planning CT scans unfortunately were no longer available. None of the patients had lobar or sublobar atelectasis. A total of 22 patients, of whom 2 had 3 tumours each, were included. In Table 1, tumour diameter, volume and position are listed for the included patients. One patient had a maximum tumour diameter of 7 cm, but was included in the study as treatment was performed anyway.

The overall results are shown in Table 2. A significant but rather weak linear correlation was present between SD_{trans} and SD_{cc} ($r = 0.42$, 95% CI = 0.04–0.69).

The patient-specific SD_{cc} was significantly larger than the patient-specific SD_{trans} ($p < 0.0001$), being consistent with the anisotropic spatial resolution of the CT scan and the lack of possibility to delineate between slices in

Table 1. Tumour characteristics

Characteristic	Value
Tumour localisation	
RUL	9
LUL	8
RML	1
RLL	3
LLL	5
Tumour diameter (cm) [Median (range)]	3.25 (0.8–7.0)
Tumour volume (cm ³) [Median (range)]	13.0 (0.3–60.4)
Pleural contact (Yes/no)	9/17
Diagnosis (patients/tumours)	
NSCLC	20/24
Lung metastasis	2/2

LLL, left lower lobe; LUL, left upper lobe; NSCLC, non-small-cell lung cancer; RLL, right lower lobe; RML, right middle lobe; RUL, right upper lobe.

the sagittal and coronal planes. We found a significant correlation between SD_{trans} and tumour volume ($r = 0.48$, 95% CI = 0.12–0.73), but not between SD_{cc} and tumour volume.

The overall mean CI_{pairwise} was 0.72, while the overall mean CI_{all} was 0.44.

The mean of the absolute tumour-specific volume deviations between largest and smallest GTV delineations was 6.4 cm³ (SD = 6.0 cm³), ranging from 0.3 to 18.3 cm³. The mean of the relative tumour-specific volume deviations between largest and smallest GTV delineations was 41% (SD = 18%), ranging from 14% to 74%. The reason for the very large relative deviations is that tumour volumes generally were small (*e.g.* the tumour with the largest relative deviation was the smallest in the study, measuring only 0.3 cm³).

Examples of tumours with large and small variations between contours in the transversal plane and CC direction, respectively, are shown in Figures 2 and 3. For the tumour with the largest interobserver delineation variation in the transversal plane (Figure 2a) there was a lack of concordance among the observers regarding an anatomical structure regarded by one of the observers as

Table 2. Overall results

SD/CI	All observers	Radiologists	Oncologists
SD_{trans}			
Mean (cm)	0.15	0.14	0.13
SD (cm)	0.08	0.07	0.08
Range (cm)	0.05–0.36	0.06–0.29	0.04–0.37
SD_{cc}			
Mean (cm)	0.26	0.27	0.20
SD (cm)	0.15	0.15	0.17
Range (cm)	0.11–0.64	0.09–0.67	0.00–0.73
CI_{all}			
Mean	0.44	0.57	0.66
SD	0.13	0.14	0.10
Range	0.21–0.71	0.32–0.83	0.46–0.85
CI_{pairwise}			
Mean	0.72	0.68	0.75
SD	0.09	0.11	0.09
Range	0.56–0.87	0.47–0.88	0.59–0.90

CC, craniocaudal; CI, concordance index; SD, standard deviation; trans, transversal plane.

Figure 2. (a) Transversal view with contours of the tumour with the largest interobserver delineation variation in the transversal plane. Standard deviation of distance to a reference contour calculated for the transversal plane (SD_{trans}) was 0.36 cm and tumour volume was 16.3 cm³. (b) Transversal views with contours of a tumour with a small interobserver delineation variation in the transversal plane. SD_{trans} was 0.07 cm and tumour volume was 1.5 cm³. Radiologists' contours are red, oncologists' contours are green and the rough positron emission tomography contour is blue.

Figure 3. (a) Sagittal views with contours of the tumour with the second largest interobserver delineation variation in the craniocaudal direction. Standard deviation of distance to a reference contour calculated for the craniocaudal direction (SD_{cc}) was 0.62 cm and tumour volume was 19.7 cm³. (b) Sagittal views with contours of two tumours in the right lung (the patient had three tumours). The caudal tumour had a small interobserver delineation variation in the craniocaudal direction, SD_{cc} being 0.11 cm. Standard deviation of distance to a reference contour calculated for the transversal plane (SD_{trans}) was 0.09 cm and tumour volume was 0.7 cm³. The cranial tumour in the right image had a SD_{cc} of 0.23 cm (SD_{trans} was 0.09 cm and tumour volume was 2.1 cm³). Radiologists' contours are red, oncologists' contours are green and the rough positron emission tomography contour is blue.

a tumour and not by the rest of the observers. For all the tumours imaged in Figure 3 there is lack of concordance between the rough PET contour and the GTV contours, stressing that PET was primarily used as a pointer and not to define the edge of the tumour.

Pleural contact or no pleural contact

For the nine tumours with pleural contact the mean tumour-specific SD_{trans} was 0.13 cm ($SD=0.07$ cm) and for the 17 tumours without pleural contact the mean tumour-specific SD_{trans} was 0.19 cm ($SD=0.08$ cm). The difference was significant ($p=0.032$). There was no significant difference in SD_{cc} , CI_{all} or $CI_{pairwise}$ between tumours with and without pleural contact.

Oncologists vs radiologists

The SD_{cc} was significantly larger for radiologists than for oncologists ($p=0.01$). There was no difference for SD_{trans} . Both calculated CIs were significantly larger for the oncologists than for the radiologists (both $p<0.0001$).

We found a significant (although weak) correlation between tumour volume and $CI_{pairwise}$ ($r=0.39$, 95% $CI=0.01$ –0.68).

Discussion

To our knowledge this study is the first to make a systematic cross-sectional analysis of delineation variation for peripheral lung tumours referred for SBRT, establishing the evidence that interobserver variation is typically small for these tumours: 0.15 cm in the transversal plane and 0.26 cm in the CC direction at the level of one SD. Nevertheless, volume deviations for GTV delineations of the same tumour of up to 18 cm³ were seen in the present study.

The interobserver variability has, however, previously been studied for other lung cancer patient groups (locally advanced disease), in a smaller set of patients and in multi-institutional settings. One of the largest systematic studies was performed by Steenbakkers et al [2], which included 22 locally advanced tumours, delineated by 11 experienced radiation oncologists from five different institutions. They analysed each anatomical region separately, and thus also the part of the tumours bordering lung tissue and the thoracic wall. They included PET information, although not dedicated FDG-PET/CT, and did not use intravenous contrast enhancement. They had a detailed protocol for delineation. They calculated a SD useable for margin calculations [11], and found a delineation variation of 0.33 cm for tumour regions bordering lung tissue and 0.37 cm for tumour regions bordering the chest wall. In another study from the same group [16], they exclusively analysed delineations of the primary tumours. They found that for the subgroup of primary tumours surrounded by lung or visceral pleura there was an interobserver variation of 0.3 cm in areas where the tumour bordered the lung or chest wall (*i.e.* despite the

subgrouping of tumours, the interobserver delineation uncertainty is twice as large as in the present study, where only tumours from patients referred for SBRT were analysed). In none of the studies is a separate interobserver delineation uncertainty for the CC direction calculated.

Previous publications have studied interobserver variation in GTV delineations for locally advanced lung cancer [1–4, 16–19] and a few have included T1–2N0M0 tumours. As expected, all studies found large overall delineation variations. This is probably mainly due to the inclusions of mediastinal tumours, but also to the lack of a delineation protocol or PET information in the majority of studies. Giraud et al [4] report CI (calculated as CI_{all}) of 0.04–0.48 in different groups of observers (also different numbers of observers in each group). Unfortunately, the majority of these studies used different methods and only a few have reported SDs of interobserver delineation variation to use for margin calculation. Our patient population was more homogeneous as all tumours were relatively small, peripheral and surrounded by lung tissue or the thoracic wall. No mediastinal tumours were included. Our observers had a wide span of experience, but all were trained in the same department. The uniform use of the PET information in this study probably also helped minimise variation. Although the delineation in this study was performed single-handedly, a reason for the relatively high conformity in this study could be that the oncologists and the radiologists in their daily clinical work co-operate in the delineation of GTVs, which may tend to produce a common view in the delineation of lung tumours. In our clinic we have found that this co-operative work promotes interdisciplinary training.

The results were influenced by the method of analysis, and the results of this and other studies may not be fully comparable. Several ratios can be used to describe the volumetric concordance: some authors have used CI [3, 4], while others have used a Dice similarity coefficient [20]. These ratios are alike but not directly comparable.

We differentiated between the transversal planes and the CC direction because this approach matches both the construction of a CT scan and the way margins are usually added for lung cancer radiotherapy in the treatment planning systems. The fact the SD_{cc} was calculated in two dimensions while the SD_{trans} was calculated in three dimensions may have impacted the results and may have contributed to the difference to some extent.

In the present study we found a significantly larger concordance among oncologists than among radiologists. This was primarily driven by uncertainties in CC direction as there was a significant difference between radiologists and oncologists in this direction and not in the transversal plane. A few studies have shown that GTV delineation differs significantly between groups [4, 21], and a single study found no difference [19]. Giraud et al [4] analysed stage I–IIIB non-small-cell lung cancer delineated by nine radiation oncologists and eight radiologists from five centres in France and Belgium. They found that the oncologists drew a significantly larger volume, but found no difference in CI (calculated as CI_{all}) for radiologists *vs* oncologists.

With regard to delineation variation, the implementation of FDG-PET/CT has the ability to reduce interobserver delineation variation [2, 6, 16, 21, 22] in locally advanced lung cancer. In many centres not only oncologists but also radiologists and specialists in nuclear medicine take part in the GTV determination and delineation process. For the small tumours treated with SBRT the large spatial resolution of PET compared with CT is a potential problem. In these cases PET primarily serves as a pointer towards the disease and is not needed to define the edges unless atelectasis is present. This is supported by a study by Fitton et al [16], providing delineation uncertainties as SDs, where the addition of PET information did not significantly impact the delineation variation for lung tumours in areas where the tumour is surrounded by lung tissue or visceral pleura.

There is presently a lack of consensus of methods, making quantitative comparisons of data difficult. The method used in the present study utilises a CoV as reference point, which is suitable for rounded tumours (such as peripheral lung tumours), but not suitable for very concavely shaped volumes, where the CoV will be placed outside the volume. The fact that the result to a high degree depends on imaging protocols and the qualifications of the observers can make the extrapolation of our results to other institutions difficult.

Image artefacts causing blurring and distortion are a major problem [23, 24], especially when imaging moving targets such as tumours in the lung, and will undoubtedly make decisions on which structures to include more difficult and uncertain for the observers.

A limitation to this study is that it is based on three-dimensional (3D) CT scans and not on four-dimensional (4D) technology. It cannot readily be assumed that these results are applicable for 4D technology or CT scans acquired using breath-hold techniques; this needs to be studied separately. However, interobserver delineation variation in the present study will probably be a reasonable estimate of the interobserver delineation variation when using the mid-ventilation phase of a 4D CT for delineation and planning [25]. Breath-hold scans may have even smaller delineation variation, as motion artefacts are less likely [23], although, given the very low uncertainties found in this work, a significant reduction is unlikely. These hypotheses obviously need confirmation from new studies.

The delineation uncertainties found for lung SBRT based on 3D CT in the present study are suited for inclusion in the probabilistic margin concept [11], which can be adapted to SBRT, as proven by Sonke and Belderbos [26].

Conclusions

This is the first study to perform a systematic cross-sectional analysis of delineation variation for peripheral lung tumours treated with SBRT and quantified as SD to use for margin computation. We found a very small interobserver delineation variation in the transversal plane (2 mm, 1 SD) and larger in the cranio-caudal direction (3 mm, 1 SD), stressing that anisotropic margins should be applied.

Conflict of interest

Stine Korreman and Per Munck af Rosenschöld have received scientific grants from Varian Medical Systems.

References

1. Vorwerk H, Beckmann G, Bremer M, Degen M, Dietl B, Fietkau R, et al. The delineation of target volumes for radiotherapy of lung cancer patients. *Radiother Oncol* 2009;91:455–60.
2. Steenbakkers RJ, Duppen JC, Fitton I, Deurloo KE, Zijp LJ, Comans EF, 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;64:435–48.
3. Kepka L, Bujko K, Garmol D, Palucki J, Zolciak-Siwinska A, Guzel-Szczepiorkowska Z, et al. Delineation variation of lymph node stations for treatment planning in lung cancer radiotherapy. *Radiother Oncol* 2007;85:450–5.
4. Giraud P, Elles S, Helfre S, De RY, Servois V, Carette MF, et al. Conformal radiotherapy for lung cancer: different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists. *Radiother Oncol* 2002;62:27–36.
5. Fitzpatrick MJ, Starkschall G, Antolak JA, Fu J, Shukla H, Keall PJ, et al. Displacement-based binning of time-dependent computed tomography image data sets. *Med Phys* 2006;33:235–46.
6. Ashamalla H, Rafla S, Parikh K, Mokhtar B, Goswami G, Kambam S, et al. The contribution of integrated PET/CT to the evolving definition of treatment volumes in radiation treatment planning in lung cancer. *Int J Radiat Oncol Biol Phys* 2005;63:1016–23.
7. Hanna GG, McAleese J, Carson KJ, Stewart DP, Cosgrove VP, Eakin RL, et al. (18)F-FDG PET-CT simulation for non-small-cell lung cancer: effect in patients already staged by PET-CT. *Int J Radiat Oncol Biol Phys* 2010;77:24–30.
8. Faria SL, Menard S, Devic S, Sirois C, Souhami L, Lisbona R, et al. Impact of FDG-PET/CT on radiotherapy volume delineation in non-small-cell lung cancer and correlation of imaging stage with pathologic findings. *Int J Radiat Oncol Biol Phys* 2008;70:1035–8.
9. Spratt DE, Diaz R, McElmurray J, Csiki I, Duggan D, Lu B, et al. Impact of FDG PET/CT on delineation of the gross tumor volume for radiation planning in non-small-cell lung cancer. *Clin Nucl Med* 2010;35:237–43.
10. Bradley J, Thorstad WL, Mutic S, Miller TR, Dehdashti F, Siegel BA, et al. Impact of FDG-PET on radiation therapy volume delineation in non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys* 2004;59:78–86.
11. van Herk M, Remeijer P, Rasch C, Lebesque JV. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. *Int J Radiat Oncol Biol Phys* 2000;47:1121–35.
12. McKenzie AL, van HM, Mijnheer B. The width of margins in radiotherapy treatment plans. *Phys Med Biol* 2000;45:3331–42.
13. Sonke JJ, Rossi M, Wolthaus J, van Herk M, Damen E, Belderbos J. Frameless stereotactic body radiotherapy for lung cancer using four-dimensional cone beam CT guidance. *Int J Radiat Oncol Biol Phys* 2009;74:567–74.
14. International Commission on Radiation Units & Measurements (ICRU). Prescribing, Recording, and Reporting Photon Beam Therapy (Report 50). Bethesda, MD; 1993.
15. Lowry L. VassarStats; Web Site for Statistical Computation cited 11 October 2010]. Available from: URL: <http://faculty.vassar.edu/lowry/VassarStats.html>
16. Fitton I, Steenbakkers RJ, Gilhuijs K, Duppen JC, Nowak PJ, van HM, 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;70:1403–7.
17. Senan S, van Sornsen de KJ, Samson M, Tankink H, Jansen P, Nowak PJ, et al. Evaluation of a target contouring protocol for 3D conformal radiotherapy in non-small cell lung cancer. *Radiother Oncol* 1999;53:247–55.
18. Van de Steene J, Linthout N, de Mey J, Vinh-Hung V, Claessens C, Noppen M, et al. Definition of gross tumor volume in lung cancer: inter-observer variability. *Radiother Oncol* 2002;62:37–49.
19. Tyng CJ, Chojniak R, Pinto PN, Borba MA, Bitencourt AG, Fogaroli RC, et al. Conformal radiotherapy for lung cancer: interobservers' variability in the definition of gross tumor volume between radiologists and radiotherapists. *Radiat Oncol* 2009;4:28.
20. Wanet M, Lee JA, Weynand B, De BM, Poncelet A, Lacroix V, et al. Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: a comparison with threshold-based approaches, CT and surgical specimens. *Radiother Oncol* 2011;98:117–25.
21. Hanna GG, Carson KJ, Lynch T, McAleese J, Cosgrove VP, Eakin RL, et al. (18)F-Fluorodeoxyglucose positron emission tomography/computed tomography-based radiotherapy target volume definition in non-small-cell lung cancer: delineation by radiation oncologists vs. joint outlining with a PET radiologist? *Int J Radiat Oncol Biol Phys* 2010;78:1040–51.
22. van Baardwijk A, Bosmans G, Boersma L, Buijsen J, Wanders S, Hochstenbag M, et al. PET-CT-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes. *Int J Radiat Oncol Biol Phys* 2007;68:771–8.
23. Persson GF, Nygaard DE, AF Rosenschold PM, Richter VI, Josipovic M, Specht L, et al. Artifacts in conventional computed tomography (CT) and Free breathing four-dimensional CT induce uncertainty in gross tumor volume determination. *Int J Radiat Oncol Biol Phys* 2011;80:1573–80.
24. Persson GF, Nygaard DE, Brink C, Jahn JW, Munck af RP, Specht L, et al. Deviations in delineated GTV caused by artefacts in 4DCT. *Radiother Oncol* 2010;96:61–6.
25. Wolthaus JW, Schneider C, Sonke JJ, van Herk M, Belderbos JS, Rossi MM, et al. Mid-ventilation CT scan construction from four-dimensional respiration-correlated CT scans for radiotherapy planning of lung cancer patients. *Int J Radiat Oncol Biol Phys* 2006;65:1560–71.
26. Sonke JJ, Belderbos J. Adaptive radiotherapy for lung cancer. *Semin Radiat Oncol* 2010;20:94–106.