Published in final edited form as:
Radiother Oncol. 2012 November ; 105(2): 167–173. doi:10.1016/j.radonc.2012.09.023.

A semiautomatic CT-based ensemble segmentation of lung tumors: comparison with oncologists' delineations and with the surgical specimen

Emmanuel Rios Velazquez[1], **Hugo J. W. L. Aerts**[1,5], **Yuhua Gu**[2], **Dmitry B. Goldgof**[3], **Dirk De Ruysscher**[1], **Andre Dekker**[1], **René Korn**[4], **Robert J. Gillies**[2], and **Philippe Lambin**[1]

[1]Department of Radiation Oncology (MAASTRO), GROW-School for Oncology and Developmental Biology, Maastricht University Medical Center (MUMC +), Maastricht, The Netherlands [2]Department of Imaging, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA [3]Department of Computer Science and Engineering, University of South Florida, Tampa, Florida USA [4]Head of Consulting Medical Imaging, Definiens AG, Munich, Germany [5]Department of Radiation Oncology and Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard University, Boston, USA

Abstract

Purpose—To assess the clinical relevance of a semiautomatic CT-based ensemble segmentation method, by comparing it to pathology and to CT/PET manual delineations by five independent radiation oncologists in non-small cell lung cancer (NSCLC).

Materials and Methods—For twenty NSCLC patients (stage Ib – IIIb) the primary tumor was delineated manually on CT/PET scans by five independent radiation oncologists and segmented using a CT based semi-automatic tool. Tumor volume and overlap fractions between manual and semiautomatic-segmented volumes were compared. All measurements were correlated with the maximal diameter on macroscopic examination of the surgical specimen. Imaging data is available on www.cancerdata.org.

Results—High overlap fractions were observed between the semi-automatically segmented volumes and the intersection (92.5 ± 9.0, mean ± SD) and union (94.2 ± 6.8) of the manual delineations. No statistically significant differences in tumor volume were observed between the semiautomatic segmentation (71.4 ± 83.2 cm^3, mean ± SD) and manual delineations (81.9 ± 94.1 cm^3; $p = 0.57$). The maximal tumor diameter of the semiautomatic-segmented tumor correlated strongly with the macroscopic diameter of the primary tumor ($r = 0.96$).

Conclusion—Semiautomatic segmentation of the primary tumor on CT demonstrated high agreement with CT/PET manual delineations and strongly correlated with the macroscopic diameter considered the "gold standard". This method may be used routinely in clinical practice and could be employed as a starting point for treatment planning, target definition in multi-center clinical trials or for high throughput data mining research. This method is particularly suitable for peripherally located tumors.

Keywords

CT; Auto-segmentation; Inter-observer variability; NSCLC delineation

Corresponding author contact information: Emmanuel Rios Velazquez, M. Sc., Department of Radiation Oncology (MAASTRO), Maastricht University Medical Center, Maastricht, The Netherlands., Maastro Clinic, Dr. Tanslaan 12, 6229 ET Maastricht, The Netherlands., Phone: 0031-88-4455666 Fax: 0031-88-4455667, emmanuel.rios@maastro.nl.

I. INTRODUCTION

Lung cancer is the deadliest type of cancer worldwide [1]. About 80 % of the lung cancer patients present advanced-stage disease (stage III and IV) and are considered inoperable due to loco-regional tumor extension, extra thoracic spread or poor physical condition at the time of diagnosis [2]. For these patients, external beam radiotherapy (RT) often combined with chemotherapy is the primary treatment modality [3].

The success of radiotherapy depends upon a good target definition and dose coverage of the target volume while limiting the radiation dose to highly radiosensitive surrounding organs. A consistent and accurate target definition is of utmost importance for accurate radiotherapy treatment planning and for treatment response evaluation. Multiple studies have reported the uncertainties and high – intra and inter – observer variability associated with target delineation in lung cancers [4–10]. Efforts have been made to reduce the observer variation for target definition, including standardized delineation protocols on CT scans and the addition of fused FDG-PET-CT information on the delineation process [11–14]. The later has diminished the inter-observer variability [15], however differences among observers are still observed for visual delineations [16]. Various PET-based methods have been developed for semiautomatic tumor delineations, ranging from simple fixed (absolute and relative) threshold based segmentations, to the more complex signal-to-background ratio and watershed clustering methods [17,18]. A few studies have compared FDG PET based automatic segmentation tools with pathological examinations [19, 20] and demonstrated their utility in reducing inter-observer variability [16, 21]. To our knowledge, no CT-based semiautomatic delineation method has been compared with both oncologists' manual delineations and with pathological examination. In practice, target volume and organs at risk are generally defined on a planning CT scan [22], which remains as the reference imaging modality in the treatment planning of non-small cell lung cancer (NSCLC). Given the observed variability, complexity and time required for target definition, a semi-automated method to accurately segment lung tumors on a CT scan would be of clinical value by providing a consistent initial target definition and would optimize the daily workflow. In this study we present a CT-based region growing method to semi-automatically segment lung tumors, that incorporates expert knowledge and it is based on the cognition network technology [23]. Our aim is to evaluate the potential clinical usefulness of a CT-based semiautomatic-segmentation method, by comparing it with CT-PET manual delineations of five independent radiation oncologists and with pathological examination of the surgical specimen.

II. Materials and methods

a. CT-PET scans

This study was approved by the local Medical Ethics Committee according to the Dutch law, and all patients provided written informed consent. Twenty consecutive patients with histologically proven non-small cell lung cancer, stage Ib – IIIb, were included in this retrospective study. All patients had undergone a diagnostic whole body PET-CT scan (Biograph, SOMATOM Sensation 16 with an ECAT ACCEL PET scanner; Siemens, Erlangen, Germany). Patients were instructed to fast at least six hours before the intravenous administration of ^{18}F-fluoro-2-deoxy-glucose (FDG) (MDS Nordion, Liège, Belgium), followed by physiologic saline (10 mL). The total injected activity of FDG was dependent on the patient weight expressed in kg: *(weight * 4) + 20* Mbq. After a period of 45 minutes, during which the patient was encouraged to rest, free-breathing PET and CT images were acquired. The CT scan was a spiral CT scan of the whole thorax with intravenous contrast. The PET images were acquired in 5-min bed positions. The CT data set was used for

attenuation correction of PET images. The complete data set was then reconstructed iteratively with a reconstruction increment of 5 mm.

b. Click and grow auto-segmentation algorithm

Pre-processing—The algorithm was developed using the Cognition Network Language (CNL) running on Definiens Developer XD with the LuTA extension [24]. CNL is, an object-based procedural computer language designed to allow automated implementation of complex, context-dependant image analysis tasks. This algorithm was designed to enable an accurate and efficient analysis of lung lesions guided by an operator. The implementation here used the original algorithm previously described [23], and was extended with a multi-seed point segmentation routine. In a first step the LuTA pre-processing was used to classify context objects like body, background, lungs and bones. They were segmented based on intensity and object size without user interaction. In a second optional step, LuTA allows an optional three-dimensional semi-automated correction of the lung boundary. The algorithm workflow is summarized in the figure 1S.

Click and grow—The third step involved the identification and segmentation of the lung lesion. The user indentified the tumor, within the segmented lung and placed a seed-point at the perceived center of the lesion. From this starting seed-point an initial seed object was automatically segmented using the LuTA region growing [23].

To improve the lesion segmentation and to reduce sensitivity towards the location of the initial seed-point, the original click-and-grow algorithm was further extended with a single click ensemble segmentation algorithm (SCES) [25]. SCES used the previously defined region, within which multiple seed points were automatically generated. Briefly, the initially segmented tumor was divided into eight regions using three perpendicular planes (xy, yz and zx), within each sub-region a seed-point was placed. Two additional seed-points were placed, one at the center of mass of the segmented tumor and one more randomly. Each seed-point was grown into a new candidate tumor region using the same LuTA region growing algorithm (multiple runs using same segmentation technique but different initial seed-point). Finally the candidate regions were merged into one consensus tumor region using a voting strategy: a voxel is classified as tumor voxel if more than half of the voxels in its 3×3×3 neighborhood window were labeled as tumor voxels in at least half of the segmented candidate tumor regions. This approach reduced inter-observer variability and operator interactions compared to the original algorithm [25].

c. GTV manual delineations

For comparison with the CT-based semiautomatic-segmentation, five observers independently carried out manual GTV delineations based on fused PET-CT images using a standard clinical delineation protocol. Briefly, the protocol included fixed window-level settings of both CT (lung W 1,700; L −300, mediastinum W 600; L 40) and PET scan (W 30,000; L 15,000) to be used for delineation [16, 26, 27]. All observers were blinded to each other delineations. The primary gross tumor volume (GTV) was defined for each patient based on combined CT and PET information. Observers were given transversal, coronal, sagittal and 3D views simultaneously. Delineations were performed on a treatment planning system (XiO; Computer Medical System, Inc., St. Louis, MO).

d. Pathology

The surgical specimen was examined according to national guidelines [28]. All patients underwent a surgical resection of their lung tumor and a standardized routine pathology examination was performed directly on the fresh specimen maintained on ice within 30 minutes after resection. Before slicing, the maximal diameter of the primary tumor was

measured by macroscopic examination. The interval time between the CT scan and the surgery or biopsy was in average 39 days (range: 7 – 112).

e. Statistical analysis

In a similarity analysis, an intersection volume (agreement between all observers) and a union volume (merging of all regions delineated by all observers) were defined and used for comparison with the semi automatically-segmented volume (Fig. 1). The overlap fraction was used to estimate the agreement between the various volumes and was defined as the volume of overlap divided by the smallest volume [29, 30]:

$$Overlap\ Fraction = \left(\frac{A \cap B}{Smallest\ volume}\right) * 100$$

where A is the semi automatically-segmented tumor and B is either the observers intersection volume or the observers union volume depending on the comparison. An overlap fraction equal to 100 indicates two perfectly matched volumes while an overlap fraction equal to zero indicates two disjoined volumes. The first overlap fraction value indicates whether the semiautomatic-segmentation method covers the common agreement (intersection volume) of the manual delineations while the second overlap fraction value indicates whether the algorithm falls within the inter-observer variability (union volume).

A volume comparison was conducted with the raw volumes expressed in cm^3. Results are summarized as the mean and standard deviation. Groups were compared with a paired Student t-test. Differences were considered to be significant when the p-value was lower than 0.05. Pearson's correlation coefficient was used to compare the maximal diameter estimates from pathology with the maximal diameter of the semi automatically-segmented volumes.

Additionally, we used the Bland-Altman analysis to evaluate the agreement between the various measurements. The Bland-Altman plot is a scatter plot that shows in the vertical axis the difference between two measures($Y - X$), against their average on the horizontal axis $\left(\frac{Y+X}{2}\right)$. Horizontal lines are superimposed in the scatter plot indicating the mean difference between the measurements and the 95% limits of agreement. If the tumor size measurements are comparable with the "gold standard", the differences in the Bland-Altman plot should be small and close to zero. A negative value means that the semi-automatic segmentation overestimates the macroscopic tumor diameter, while a positive value indicates that the semi-automatic segmentation underestimates the gold standard.

All data are expressed as mean ± SD. All the analyses were performed in Matlab 2010b (The MathWorks Inc, Natick, MA, USA).

III. Results

To evaluate clinical validity, semiautomatic-segmentation of the primary tumor on CT was compared with CT/PET manual delineations of five independent observers.

A similarity analysis was performed and the overlap fraction was calculated to estimate the agreement between the manually contoured and SCES volumes. The overlap fractions between the semi automatically-segmented volumes and the observers union and intersection are shown in figure 2. High overlap fractions were obtained with the observers' intersection (92.5 ± 9.0) and the observers union (94.2 ± 6.8).

The raw semi automatically-segmented and manual volumes are reported in Table 1, mean and standard deviation of the manual volumes are shown as well. No statistical differences were observed in tumor volume between the SCES volumes (71.4 ± 83.2 cm^3) and manual delineations (81.9 ± 94.1 cm^3; $p = 0.57$). In the majority of the cases the semi automatically-segmented volumes fell within the observers' variability, i.e. 75% of the cases were included in the mean ± 1 SD range of the manually delineated volumes. Three cases fell off the observers' variability, two of which were centrally located tumors and one peripherally located. For visual comparison, representative examples of both the semi-automatically segmented volumes and manual delineations are shown in figure 3.

To further evaluate its usability, the click and grow semiautomatic-segmentation algorithm was compared with macroscopic examination of the surgical specimen. A strong correlation was found between the maximal diameter of the SCES volumes and the macroscopic diameter of the primary tumors (Pearson correlation coefficient, 0.96) (Fig 4). The correlation of maximal diameters on manual CT/PET delineations with the pathology examination ranged from 0.88 to 0.96 (0.93 ± 0.02) for different observers.

Tumor diameters ranged from 1.8 to 9.0 cm on pathological examination (4.2 ± 1.9), from 1.7 to 9.8 (5.1 ± 1.7) on manual CT/PET delineations and from 2.0 to 9.1 (4.9 ± 2.2) on CT semi automatically-segmented tumors.

Comparative Bland-Altman plots are shown in figure 5, the CT semiautomatic-segmentation algorithm slightly overestimated the maximal tumor diameter found at macroscopic examination, indicated by a mean difference of −0.46 cm, (95 % CI, −1.64 – 0.70 cm). However, for manual CT/PET delineations, the differences were slightly larger (mean difference −0.80 cm; 95% CI, −2.42 – 0.82 cm) compared to semi automatically-segmented volumes. For the majority of the patients, the difference between the macroscopic diameter and the diameters on the semi automatically-segmented volumes was within 1 cm.

IV. Discussion

Target definition remains highly dependent on human interpretation of visual imaging information, making it error prone and subjective. Uncertainties associated with target definition have been largely reported, and especially for lung cancer, high intra- and inter-observer variability have been observed [4–10]. In practice, human interaction, regarding CT/PET visual interpretation of imaging information and tuning of parameters in complex auto-segmentation algorithms still remains as the largest source of uncertainty for target definition [5, 7].

To our knowledge, this is the first study that compares a semi automated method to segment lung tumors on CT images, with multiple manual delineations used routinely in the clinic and furthermore compared to pathology. CT-Based semiautomatic ensemble segmentation of lung tumors showed high agreement with radiation oncologist's manual CT/PET delineations and correlated with pathological tumor measurements (Pearson's correlation, 0.96).

Multiple methods have been developed aiming to improve target definition [17–20]. For example, in the case of PET imaging, the simplest method uses an absolute threshold of the standardized uptake value. Single threshold methods are subject to considerable variation due to heterogeneities in tumor size and FDG-uptake and lack of standardization in PET acquisition settings [31]. More sophisticated segmentation algorithms such as the individualized threshold based on the source-to-background ratio [16, 32], the gradient based delineation or the watershed clustering have been proposed, and are generally preferred over single threshold methods [18, 20, 33]. These studies showed the large

differences on the resulting volumes depending on the delineation method. Recently, Wanet et al. compared different automatic delineation methods, including the gradient based method, the source-to-background based method and fixed thresholds at 40% and 50% of the SUV_{max}, with manually delineated contours on the macroscopic specimen and on CT images [20]. Although this method was compared with a three-dimensional reconstruction of the macroscopic specimen, only ten patients were evaluated in the study. Despite these efforts, a consensus over the most reliable method for automatic delineation of the GTV based on PET is lacking.

CT remains as the reference imaging modality for the treatment planning in NSCLC [22], it is widely available and the image acquisition protocols are better standardized across institutions compared to PET. Thus, we believe a CT based semiautomatic method to segment lung tumors that limits user interaction has clinical value.

In this regard, automatic methods to interpret chest CT images have largely focused on the early detection of lung nodules [34, 35], on the differential diagnosis of malignant versus benign nodules [36] and in the measurement of nodule or tumor size as treatment response criteria [37, 38]. This study however, concentrates on the definition of locally advanced stage primary tumors, with the exception of two stage Ib patients, which tend to be large, irregular masses often adjacent to other anatomical structures, with the minimal user guidance. Other studies have used supervised approaches using different machine learning techniques to classify lung tissues [39, 40], however these techniques are trained with manually delineated regions of interest, rather than performing automatic detection and were not adequately validated. A general disadvantage of the supervised approach is that it requires beforehand labeled samples used for training of the algorithm.

Kaker et al. proposed an automatic method to classify lung lesions and healthy tissue on CT images, being able to classify the tissue in twelve different categories, differing mainly based on location (i. e. left lung upper, left lung lower) rather than on tissue type [41]. Although fairly accurate classification values were reported they failed to validate their approach since the segmentations were not adequately compared with a ground truth estimation. Furthermore, they artificially generated 500 samples by under or over sampling the original data to develop and evaluate their method, which questions the validity of the testing data.

Our study allowed the possibility of comparing our results with examination of the surgical specimen, by including surgical patients. However, this can also bias the algorithm performance towards peripherally located tumors. In fact, in 75% of the cases, of which 40% were centrally located, semiautomatic-segmentation of tumors on CT images proved to be adequate and could probably be further extended to late stage, inoperable patients treated with radiotherapy. However, in 25% of the cases, the algorithm showed reduced overlap with the observer's delineation. Two of the cases were centrally located tumors, these cases had large central cavities that were covered by the observer's delineation (shown in Figure 4), but that the algorithm judges as not being part of the tumor. This largely explains the apparent mismatch between the observer's delineations and the semi auto-segmented volumes. Furthermore, the centrally located tumors also displayed larger observer's variability. In other tumors with central location, the algorithm showed overlap fractions of approx. 95%.

An additional case with lower overlap fractions was a small isolated tumor in the upper lobe of the left lung. In this case the algorithm overestimated the tumor extension, presumably by extending the tumor voxels to an adjacent bronchiole. In these situations, supervision is warranted.

Our results showed a high correlation with the maximal diameter of the surgical specimen measured on macroscopic examination. However, there are limitations intrinsic to the method employed to determine the pathological tumor diameter that should be addressed. To date only a few studies have validated auto-delineation methods with the gold standard of pathology [16, 20, 42]. Daisne et al. and Stroom et al. have proposed techniques to obtain a three-dimensional digital reconstruction of the pathological specimen, by fixating, slicing and photographing the surgical specimen under controlled conditions [42, 43]. In our study, the maximal diameter on pathology was determined with a ruler in one dimension, a fairly simple measurement prone to tumor shrinkage and deformation. Measurements were performed before fixation which reduces the influence of tumor shrinkage. In future work, three-dimensional assessment of the pathological specimen is warranted. Of importance is the time interval between the imaging study and the date of surgery or biopsy, which was in average 39 days (range: 7–112 days). This time interval could impact the results in case of tumor growth. This relatively large time span could not be avoided as those were the treatment schedules. However, if large changes in tumor growth occurred between the day of the imaging study and the day of the surgical intervention, these would have been observed in the Bland-Altman analysis. This analysis showed that the difference between the macroscopic diameter and the diameters on the semi automatically-segmented volumes was within 1 cm in 85 % of the cases. We believe this has clinical value in terms that the semi-auto segmented volumes fairly compare to those delineated by the clinical experts and the proposed method agrees with the gold standard at least as good as the medical experts as demonstrated by the Bland-Altman plot.

Because of the retrospective nature of this study, we lacked thin-slice reconstruction CT images; slice thickness in our data was 5 mm. Thin-slice CT images could enhance the ability to accurately determine the tumor volume.

Besides reducing inter-observer variability, semi-automated segmentation of tumors has proven useful to reduce the target delineation time in other tumor sites [44–46]. The time needed to manually delineate the tumors was not recorded, however, in a multi-centric study Steenbakkers reported a mean delineation time of 16 min (SD 10 min) on CT scans and of 12 min (SD 8 min) on CT/PET scans [15]. The time needed to segment the lung tumor with the semiautomatic software, including importing and loading the CT data, pre-processing, initial seed-point definition and ensemble segmentation was in average 12.02 min (SD 0.4 min). Finally, this method could be relevant for adaptive radiotherapy, where treatment plans are modified under restricted time slots, and in the first instance could be used as a pre-delineated structure that the clinical expert could modify. Since this algorithm fairly compared with the standardized delineations that are the current clinical practice, we view this algorithm as an approximation of the target volume that could facilitate the target definition step in radiotherapy treatment planning, but not as a replacement of the medical experts.

To conclude, this method could be employed for target definition routinely on clinical practice, as an approximation that can be refined by the medical experts, and in high-throughput data mining research, based on the overlap with manual delineations and its correlation with the pathological examination, in particular for peripherally located tumors. Further validation will require a multi-center investigation.

Acknowledgments

This study was supported by the Dutch Cancer Society (KWF UM-2010-4776) and the National Institutes of Health (NIH-USA U01 CA 143062-01, Radiomics of NSCLC). The authors wish to acknowledge Definiens AG, for the algorithm development and support.

VI. References

1. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin. 60:277–300. [PubMed: 20610543]
2. Auperin A, Le Pechoux C, Rolland E, et al. Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small-cell lung cancer. J Clin Oncol. 28:2181–2190. [PubMed: 20351327]
3. Jett JR, Schild SE, Keith RL, Kesler KA. Treatment of non-small cell lung cancer, stage IIIB: ACCP evidence-based clinical practice guidelines (2nd edition). Chest. 2007; 132:266S–276S. [PubMed: 17873173]
4. Bradley J, Bae K, Choi N, et al. A phase II comparative study of gross tumor volume definition with or without PET/CT fusion in dosimetric planning for non-small-cell lung cancer (NSCLC): primary analysis of Radiation Therapy Oncology Group (RTOG) 0515. Int J Radiat Oncol Biol Phys. 2012; 82:435–441. e431. [PubMed: 21075551]
5. 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. Radiother Oncol. 2002; 62:27–36. [PubMed: 11830310]
6. Grabarz D, Panzarella T, Bezjak A, McLean M, Elder C, Wong RK. Quantifying interobserver variation in target definition in palliative radiotherapy. Int J Radiat Oncol Biol Phys. 2011; 80:1498–1504. [PubMed: 20656422]
7. Greco C, Rosenzweig K, Cascini GL, Tamburrini O. Current status of PET/CT for tumour volume definition in radiotherapy treatment planning for non-small cell lung cancer (NSCLC). Lung Cancer. 2007; 57:125–134. [PubMed: 17478008]
8. Steenbakkers RJ, Duppen JC, Fitton I, et al. Observer variation in target volume delineation of lung cancer related to radiation oncologist-computer interaction: a 'Big Brother' evaluation. Radiother Oncol. 2005; 77:182–190. [PubMed: 16256231]
9. Van de Steene J, Linthout N, de Mey J, et al. Definition of gross tumor volume in lung cancer: inter-observer variability. Radiother Oncol. 2002; 62:37–49. [PubMed: 11830311]
10. Vorwerk H, Beckmann G, Bremer M, et al. The delineation of target volumes for radiotherapy of lung cancer patients. Radiotherapy and Oncology. 2009; 91:455–460. [PubMed: 19339069]
11. Senan S, van Sornsen de Koste J, Samson M, et al. Evaluation of a target contouring protocol for 3D conformal radiotherapy in non-small cell lung cancer. Radiother Oncol. 1999; 53:247–255. [PubMed: 10660205]
12. Bowden P, Fisher R, Mac Manus M, et al. Measurement of lung tumor volumes using three-dimensional computer planning software. Int J Radiat Oncol Biol Phys. 2002; 53:566–573. [PubMed: 12062598]
13. Caldwell CB, Mah K, Ung YC, et al. Observer variation in contouring gross tumor volume in patients with poorly defined non-small-cell lung tumors on CT: the impact of 18FDG-hybrid PET fusion. Int J Radiat Oncol Biol Phys. 2001; 51:923–931. [PubMed: 11704312]
14. Buijsen J, van den Bogaard J, van der Weide H, et al. FDG-PET-CT reduces the interobserver variability in rectal tumor delineation. Radiotherapy and Oncology. 2012; 102:371–376. [PubMed: 22280807]
15. 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]
16. van Baardwijk A, Bosmans G, Boersma L, 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; 68:771–778. [PubMed: 17398018]
17. Lee JA. Segmentation of positron emission tomography images: some recommendations for target delineation in radiation oncology. Radiother Oncol. 2010; 96:302–307. [PubMed: 20708286]
18. Nestle U, Kremp S, Schaefer-Schuler A, et al. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer. J Nucl Med. 2005; 46:1342–1348. [PubMed: 16085592]

19. Wu K, Ung YC, Hornby J, et al. PET CT thresholds for radiotherapy target definition in non-small-cell lung cancer: how close are we to the pathologic findings? Int J Radiat Oncol Biol Phys. 2009; 77:699–706. [PubMed: 19836163]
20. Wanet M, Lee JA, Weynand B, 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. 2010; 98:117–125. [PubMed: 21074882]
21. van Loon J, van Baardwijk A, Boersma L, Ollers M, Lambin P, De Ruysscher D. Therapeutic implications of molecular imaging with PET in the combined modality treatment of lung cancer. Cancer Treat Rev. 2011; 37:331–343. [PubMed: 21320756]
22. Sonke JJ, Belderbos J. Adaptive radiotherapy for lung cancer. Semin Radiat Oncol. 2010; 20:94–106. [PubMed: 20219547]
23. Bendtsen C, Korn MKR, Mozley PD, Schmidt G, Binnig G. X-Ray Computed Tomography: Semiautomated Volumetric Analysis of Late-Stage Lung Tumors as a Basis for Response Assessments. International Journal of Biomedical Imaging. 2011; 2011:11.
24. Athelogou, M.; GS; Schaepe, A.; Baatz, M.; Binnig, G. Imaging Cellular and Molecular Biological Functions. New York, NY, USA: Springer; 2007. Definiens cognition network technology—a novel multimodal image analysis technique for automatic identification and quantification of biological image contents.
25. Gu, Y.; Kumar, V.; Hall, L., et al. Automated delineation of lung tumors from CT images: method and evaluation. World Molecular Imaging Congress; San Diego. 2011. p. 373
26. van Baardwijk A, Bosmans G, Boersma L, et al. Individualized radical radiotherapy of non-small-cell lung cancer based on normal tissue dose constraints: a feasibility study. Int J Radiat Oncol Biol Phys. 2008; 71:1394–1401. [PubMed: 18258380]
27. van Baardwijk A, Wanders S, Boersma L, et al. Mature results of an individualized radiation dose prescription study based on normal tissue constraints in stages I to III non-small-cell lung cancer. J Clin Oncol. 2010; 28:1380–1386. [PubMed: 20142596]
28. Richtlijn Niet-kleincellig longcarcinoom: stadiering en behandeling. Vereniging van Integrale Kankercentra; 2004. http://www.cbo.nl/thema/Richtlijnen/Overzicht-richtlijnen/Oncologie/
29. Aerts HJ, van Baardwijk AA, Petit SF, et al. Identification of residual metabolic-active areas within individual NSCLC tumours using a pre-radiotherapy (18)Fluorodeoxyglucose-PET-CT scan. Radiother Oncol. 2009; 91:386–392. [PubMed: 19329207]
30. Hanna GG, Hounsell AR, O'Sullivan JM. Geometrical analysis of radiotherapy target volume delineation: a systematic review of reported comparison methods. Clin Oncol (R Coll Radiol). 2010; 22:515–525. [PubMed: 20554168]
31. Nestle U, Kremp S, Grosu AL. Practical integration of [18F]-FDG-PET and PET-CT in the planning of radiotherapy for non-small cell lung cancer (NSCLC): the technical basis, ICRU-target volumes, problems, perspectives. Radiother Oncol. 2006; 81:209–225. [PubMed: 17064802]
32. Daisne JF, Sibomana M, Bol A, Doumont T, Lonneux M, Gregoire V. Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. Radiother Oncol. 2003; 69:247–250. [PubMed: 14644483]
33. Hatt M, Cheze le Rest C, Descourt P, et al. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. Int J Radiat Oncol Biol Phys. 2010; 77:301–308. [PubMed: 20116934]
34. Dehmeshki J, Amin H, Valdivieso M, Ye X. Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach. IEEE Trans Med Imaging. 2008; 27:467–480. [PubMed: 18390344]
35. Armato SG 3rd, Giger ML, MacMahon H. Automated detection of lung nodules in CT scans: preliminary results. Med Phys. 2001; 28:1552–1561. [PubMed: 11548926]
36. McNitt-Gray MF, Hart EM, Wyckoff N, Sayre JW, Goldin JG, Aberle DR. A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: preliminary results. Med Phys. 1999; 26:880–888. [PubMed: 10436888]
37. Nishino M, Guo M, Jackman DM, et al. CT tumor volume measurement in advanced non-small-cell lung cancer: Performance characteristics of an emerging clinical tool. Acad Radiol. 2011; 18:54–62. [PubMed: 21036632]

38. Yankelevitz DF, Reeves AP, Kostis WJ, Zhao B, Henschke CI. Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. Radiology. 2000; 217:251–256. [PubMed: 11012453]

39. Uchiyama Y, Katsuragawa S, Abe H, et al. Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography. Med Phys. 2003; 30:2440–2454. [PubMed: 14528966]

40. Ye X, Beddoe G, Slabaugh G. Automatic Graph Cut Segmentation of Lesions in CT Using Mean Shift Superpixels. Int J Biomed Imaging. 2010:983963. [PubMed: 21052498]

41. Kakar M, Olsen DR. Automatic segmentation and recognition of lungs and lesion from CT scans of thorax. Comput Med Imaging Graph. 2009; 33:72–82. [PubMed: 19059759]

42. Stroom J, Blaauwgeers H, van Baardwijk A, et al. Feasibility of pathology-correlated lung imaging for accurate target definition of lung tumors. Int J Radiat Oncol Biol Phys. 2007; 69:267–275. [PubMed: 17707281]

43. Daisne JF, Duprez T, Weynand B, et al. Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. Radiology. 2004; 233:93–100. [PubMed: 15317953]

44. Strassmann G, Abdellaoui S, Richter D, et al. Atlas-based semiautomatic target volume definition (CTV) for head-and-neck tumors. Int J Radiat Oncol Biol Phys. 2010; 78:1270–1276. [PubMed: 20630666]

45. Anders LC, Stieler F, Siebenlist K, Schäfer J, Lohr F, Wenz F. Performance of an atlas-based autosegmentation software for delineation of target volumes for radiotherapy of breast and anorectal cancer. Radiotherapy and Oncology. 2012; 102:68–73. [PubMed: 21962822]

46. Isambert, Al; Dhermain, Fdr; Bidault, Fo, et al. Evaluation of an atlas-based automatic segmentation software for the delineation of brain organs at risk in a radiation therapy clinical context. Radiotherapy and Oncology. 2008; 87:93–99. [PubMed: 18155791]

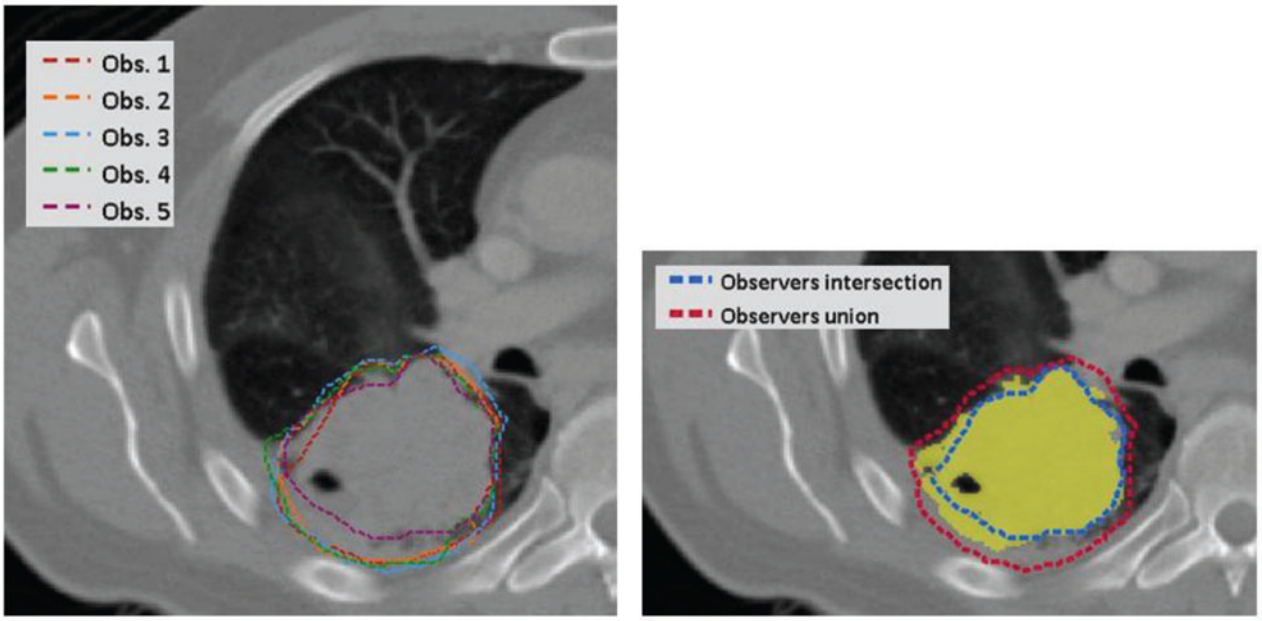

Figure 1.
The image in the left side shows the variability observed for CT/PET manual delineations. To summarize inter-observer variability the observers intersection (common agreement) and observers union (sum of all delineated areas) were defined and compared with the CT semi auto-segmentation method (yellow color wash) in the right panel. These images correspond to patient 12.

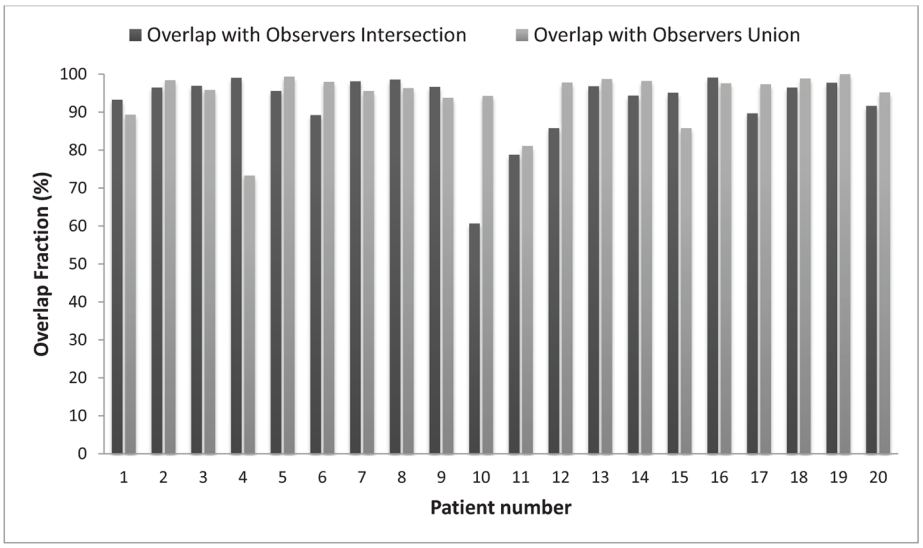

Figure 2.
Overlap fractions between the semi auto-segmented volumes and observers' intersection (agreement between all observers) and union (merging of all regions delineated by all observers) volumes. An overlap fraction equal to 100 indicates two perfectly matched volumes while an overlap fraction equal to zero indicates two disjoined volumes.

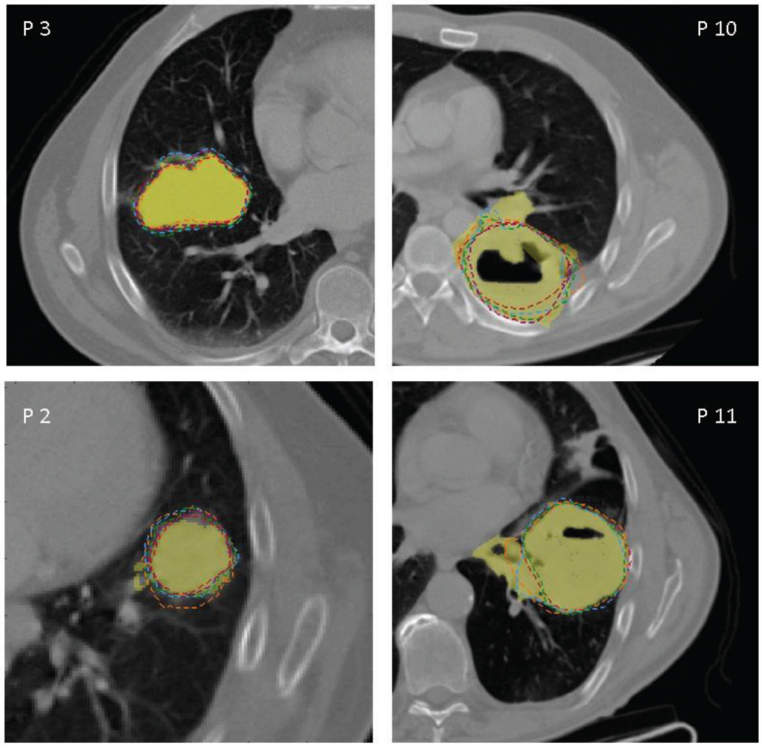

Figure 3.
Representative CT images of NSCLC patients. Lung tumors were segmented using a click and grow ensemble segmentation algorithm (yellow solid color wash) and manually delineated by five independent observers (color dotted lines).

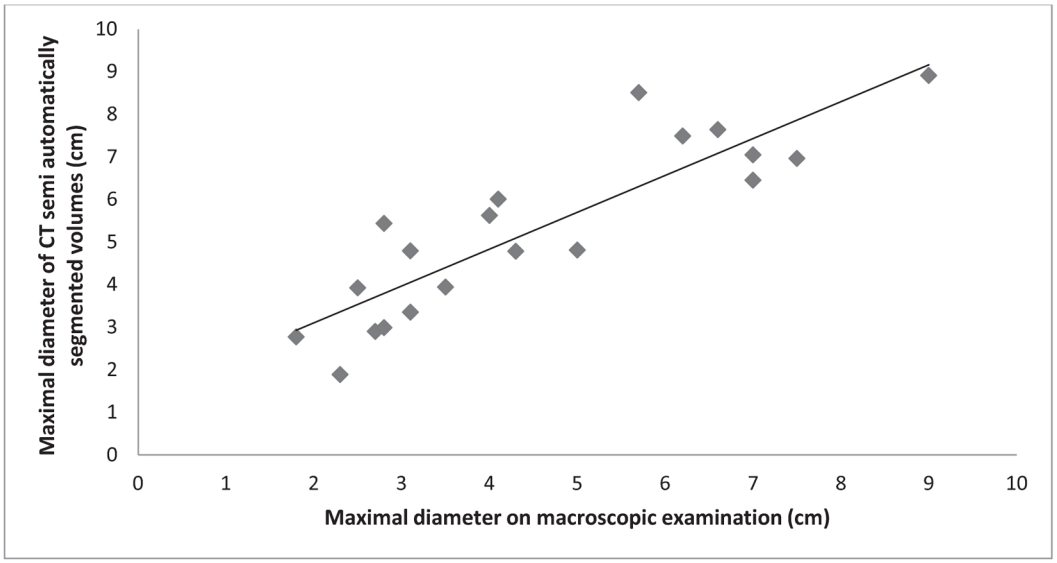

Figure 4.
Maximal diameter of the primary tumor determined by the CT semi auto-segmentation algorithm compared with the tumor maximal diameter on macroscopic examination of the surgical specimen. Pearson correlation coefficient was 0.96.

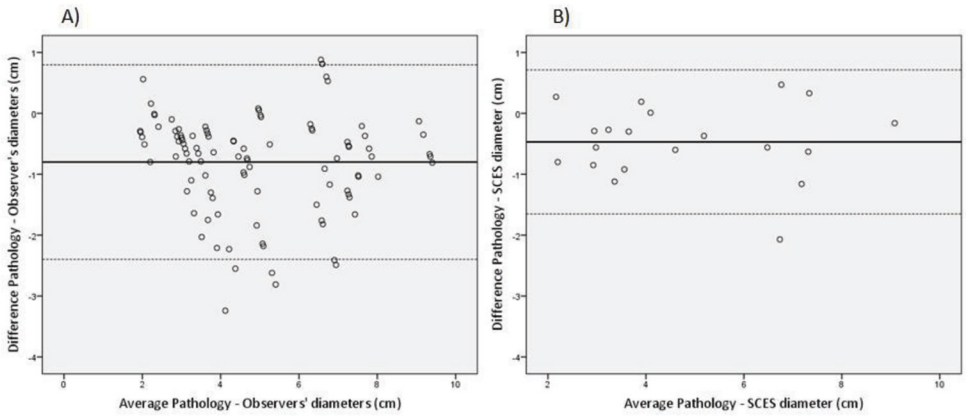

Figure 5.

Bland Altman plots showing the discrepancies of CT/PET manual delineations (A) and CT semi auto-segmentations (B) with the 'golden truth' maximal diameter on pathology. The dotted lines represent the confidence intervals around the mean difference (solid line).

Radiother Oncol. Author manuscript; available in PMC 2013 November 15.

Table 1

Raw volumes in cm^3 as determined by CT/PET manual delineations and compared with the CT semi auto-segmentation method.

Patient No.	Tumor stage	Location	Observer 1	Observer 2	Observer 3	Observer 4	Observer 5	Mean observers	SD observers	Semi auto-segmentation
1	IIIa	Central	69.8	79.4	103.9	80.3	69.5	**80.6**	**14.0**	83.9
2	IIIa	Peripheral	8.0	17.6	13.1	11.9	8.5	**11.8**	**3.9**	9.0
3	IIIa	Central	355.0	320.4	380.4	353.1	309.5	**343.7**	**28.6**	334.6
4	IIIa	Peripheral	3.1	4.9	4.8	4.2	3.2	**4.1**	**0.9**	5.8
5	IIIa	Peripheral	10.6	16.1	21.2	15.9	10.2	**14.8**	**4.5**	10.4
6	IIIa	Central	188.1	219.0	275.2	206.8	150.7	**208.0**	**45.6**	163.9
7	Ib	Peripheral	46.1	59.2	57.6	57.1	48.4	**53.7**	**6.0**	50.1
8	IIIa	Peripheral	18.0	35.5	43.2	25.5	16.8	**27.8**	**11.4**	26.2
9	IIIa	Central	29.8	31.8	37.9	34.5	28.5	**32.5**	**3.8**	31.6
10	IIIa	Central	196.4	266.8	215.4	259.6	175.1	**222.7**	**39.7**	146.1
11	IIIa	Central	151.2	184.2	184.6	164.0	144.3	**165.7**	**18.5**	150.7
12	IIIa	Peripheral	199.1	193.5	193.4	194.9	205.1	**197.2**	**5.0**	158.2
13	IIIa	Peripheral	55.9	77.9	65.6	68.5	56.8	**65.0**	**9.1**	58.0
14	IIIa	Peripheral	6.7	9.4	12.2	9.3	9.1	**9.3**	**1.9**	7.9
15	IIIa	Central	18.2	21.9	12.5	14.5	12.5	**15.9**	**4.1**	19.6
16	IIIa	Peripheral	38.5	40.9	24.2	36.3	34.3	**34.8**	**6.5**	36.7
17	IIIa	Central	31.0	37.1	39.9	36.6	31.7	**35.3**	**3.8**	29.5
18	Ib	Peripheral	82.2	101.9	98.9	97.9	92.4	**94.7**	**7.8**	88.2
19	IIb	Peripheral	5.7	5.1	1.7	4.8	4.0	**4.3**	**1.6**	2.6
20	IIIa	Peripheral	12.2	14.0	30.3	19.1	9.8	**17.1**	**8.1**	14.5