DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Volumetric CT-based segmentation of NSCLC using 3D-Slicer

Citation	Velazquez, E. R., C. Parmar, M. Jermoumi, R. H. Mak, A. van Baardwijk, F. M. Fennessy, J. H. Lewis, et al. 2013. "Volumetric CT-based segmentation of NSCLC using 3D-Slicer." Scientific Reports 3 (1): 3529. doi:10.1038/srep03529. http://dx.doi.org/10.1038/srep03529.
Published Version	doi:10.1038/srep03529
Accessed	April 17, 2018 4:43:06 PM EDT
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:11879380
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)

Volumetric CT-based segmentation of NSCLC using 3D-Slicer

Emmanuel Rios Velazquez[1,2]*, Chintan Parmar[1,2]*, Mohammed Jermoumi[1,3], Raymond H. Mak[1], Angela van Baardwijk[2], Fiona M. Fennessy[5], John H. Lewis[1], Dirk De Ruysscher[4], Ron Kikinis[5], Philippe Lambin[2] & Hugo J. W. L. Aerts[1,2]

[1]Department of Radiation Oncology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, [2]Department of Radiation Oncology (MAASTRO), GROW Research Institute, Maastricht University, Maastricht, the Netherlands, [3]University of Massachusetts, Lowell, MA, USA, [4]Radiation Oncology, University Hospitals Leuven/KU Leuven, Leuven, Belgium, [5]Department of Radiology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.

Accurate volumetric assessment in non-small cell lung cancer (NSCLC) is critical for adequately informing treatments. In this study we assessed the clinical relevance of a semiautomatic computed tomography (CT)-based segmentation method using the competitive region-growing based algorithm, implemented in the free and public available 3D-Slicer software platform. We compared the 3D-Slicer segmented volumes by three independent observers, who segmented the primary tumour of 20 NSCLC patients twice, to manual slice-by-slice delineations of five physicians. Furthermore, we compared all tumour contours to the macroscopic diameter of the tumour in pathology, considered as the "gold standard". The 3D-Slicer segmented volumes demonstrated high agreement (overlap fractions > 0.90), lower volume variability ($p = 0.0003$) and smaller uncertainty areas ($p = 0.0002$), compared to manual slice-by-slice delineations. Furthermore, 3D-Slicer segmentations showed a strong correlation to pathology (r = 0.89, 95%CI, 0.81–0.94). Our results show that semiautomatic 3D-Slicer segmentations can be used for accurate contouring and are more stable than manual delineations. Therefore, 3D-Slicer can be employed as a starting point for treatment decisions or for high-throughput data mining research, such as Radiomics, where manual delineating often represent a time-consuming bottleneck.

Lung cancer is a disease that affects about 1.6 million individuals worldwide every year[1]. Non-small cell lung cancer (NSCLC) accounts for 85% of all lung cancer cases and it is characterized by poor prognosis and low survival rates, due to high incidence of loco-regional and distant recurrences[2].

In lung cancer, tumour delineation is critical for accurate volumetric assessment to evaluate response to therapy, which can inform treatment decisions. However, tumour delineation can be a source of uncertainty, since typically, the tumour delineation process involves an experienced physician, interpreting and manually contouring computed tomography (CT) alone or combined with Fluorodeoxyglucose (FDG) - positron emission tomography (PET) imaging, on a slice-by-slice basis[3–6]. Despite efforts in standardization of CT or FDG-PET-CT image acquisition and standardized guidelines for tumour delineation, definition of lung tumours remains prone to inter-observer variability and is time consuming[6–9].

To reduce these problems, a number of CT or FDG-PET based semi-automatic methods have been investigated, that aim to provide equivalent segmentations to those delineated manually by physicians, or to provide a starting point for the manual delineation process, thereby reducing the overall required time. The various segmentation methods, that range from simple threshold based methods to complex level set, watershed, or region growing-context based methods, have been compared to manual delineations provided by physicians and compared to the pathological measurements of tumour size, with varying success rates[10–16]. However, the application of these methods is limited, often due to accessibility of the method within the clinical delineation process.

In this study we evaluated the utility of the GrowCut algorithm to segment lung tumours, implemented in 3D-Slicer – a free open source software platform for biomedical research[17]. This cellular automaton-based algorithm performs automatic tumour segmentation after drawing boundaries within the image volume. It provides an alternative to the manual slice-by-slice segmentation process and is found to be significantly faster and less user intensive[17]. Our hypothesis is that 3D-Slicer contours are more stable for inter-observer variation compared to

manual contouring. To evaluate the accuracy of the 3D-Slicer segmentations, three independent observers segmented 20 NSCLC patients twice using 3D-Slicer. We compared these six 3D-Slicer segmentations to manual delineations provided by five physicians. Furthermore, the segmented volumes were compared with the maximum diameter measured from the tumour after resection, considered as the gold standard. Because 3D-Slicer is publicly available and easily accessible by download, its application in NSCLC could be useful for the clinical investigations where tumour contours are necessary for assessing therapy response, therapy planning, or in high-throughput data mining research of medical imaging in clinical oncology (Radiomics)[18–21].

Results

Clinical reliability of the 3D slicer's semi-automatic segmentations was measured in terms of its agreement with the CT/PET manual tumour delineations of five independent observers and with pathological measurements after surgery. To quantify the agreement between the manual and 3D-Slicer segmentations, we performed an uncertainty analysis. The uncertainty region was defined as the region that varied between the segmentations of the different observers. In figure 1, the uncertainty region of five manual and six 3D-Slicer segmentations (three observers segmented twice with different seed-point initialisation) is illustrated. This example shows that the uncertainty region is larger for manual delineations compared to 3D-Slicer.

Overlap fractions. To examine the spatial agreement of the manual and 3D-Slicer contours, Overlap Fractions (OF) were calculated. OFs were computed between each of the six 3D-Slicer segmentations with the uncertainty region of the manual delineations. The intersection is defined as the inner boundary of the uncertainty region (i.e. the region that all manual observers delineated), and the union as the outer boundary of the uncertainty region (i.e. the region at least one of the manual observers delineated). High OFs were observed with the observers' intersection (mean ± SD: 94.3 ± 4.4%, range: 76.8–99.8) and union (mean ± SD: 97.2 ± 5.1%; range: 72.6–100) [See figure 2]. In the Supplementary Figure S1, a heat map depicting the overlap fractions for each patient between the GrowCut segmentations and manual delineations' union and intersection are shown. The results demonstrate a high spatial agreement of the manual and 3D-Slicer segmentations.

Uncertainty regions. To investigate the robustness of 3D-Slicer segmentations we compared its uncertainty region against the manual uncertainty region [Figure 1]. The analysis showed that the uncertainty region, defined as the difference between uncertainty region inner and outer boundaries, was smaller for the 3D-Slicer segmentations [See Figure 3A]. Manual delineations had significantly larger uncertainty areas compared to 3D-Slicer segmentations (Wilcoxon test $p = 0.0002$).
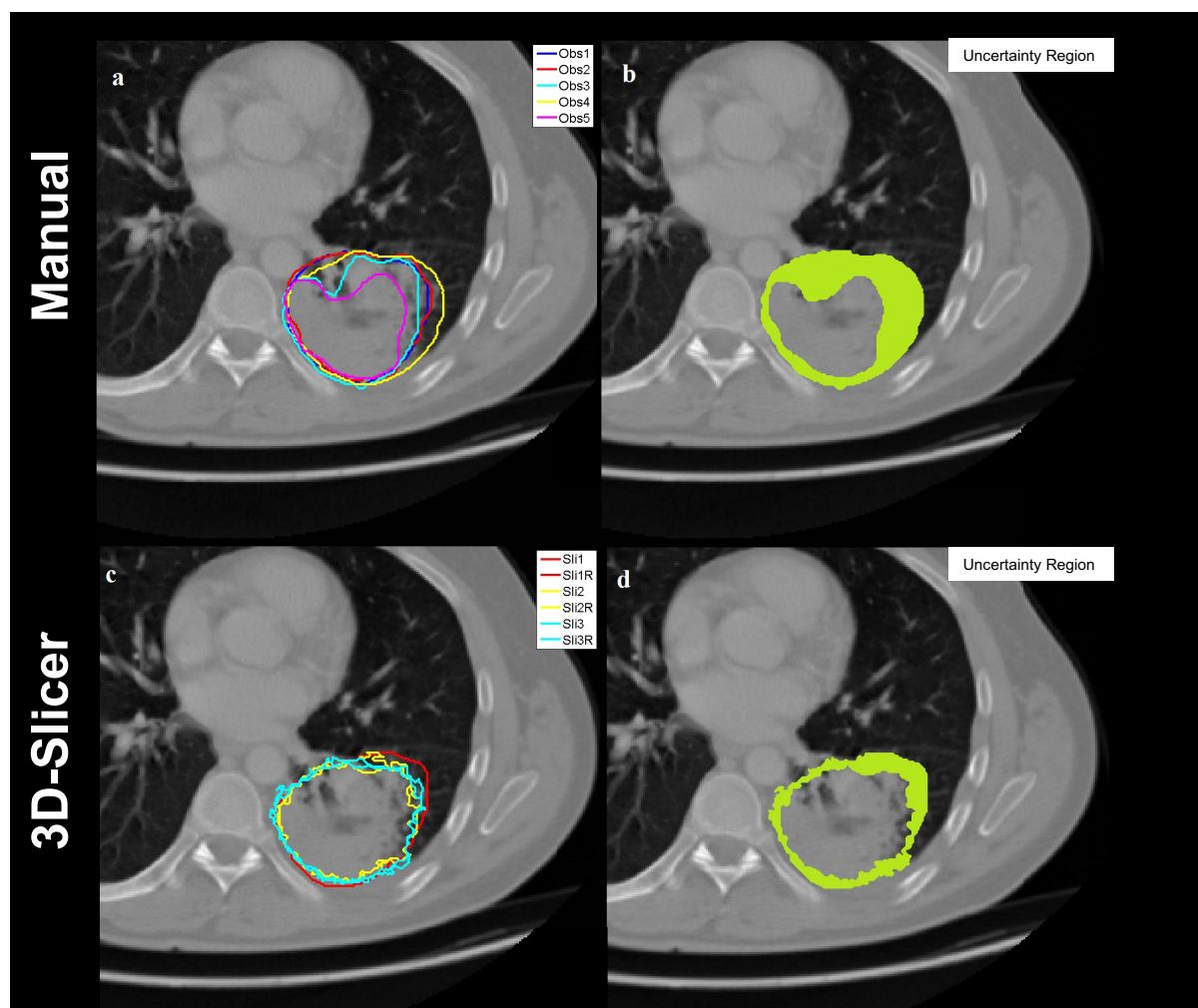

Figure 1 | Segmentation uncertainty. Left: representative example showing differences in CT/PET manual delineations (top) and 3D-Slicer segmentations (bottom). Right: This variability is quantified with the uncertainty region, defined as the difference between the observers' agreement and observers' union (highlighted in green). The smaller the uncertainty region is, the lower the variability among multiple contours.
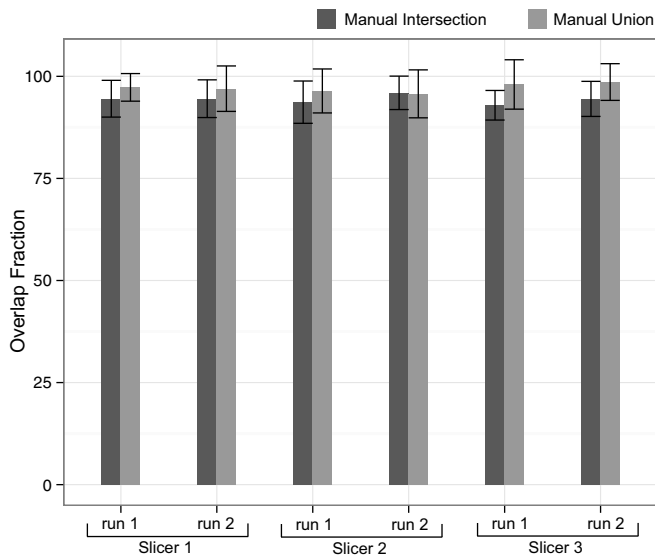
Figure 2 | Overlap fractions between the 3D-Slicer segmented volumes and the observers' intersection and union volumes. High overlap fraction indicates high agreement (spatial overlap) between volumes.

Segmented volumes. We then investigated the volumes of the segmentations. There was a high agreement between the volumes of the manual and 3D-Slicer contours, as we found no statistically significant difference between the volumes of the five manual delineations (82.03 ± 94.31 cm^3) and six 3D-Slicer (72.27 ± 86.62 cm^3, mean ± SD) segmentations, using Kruskal–Wallis one-way analysis of variance ($p = 0.98$). Figure 3B, displays the tumour volume variability, for both manual and 3D-Slicer for all patients. In 17 cases (85%), the volume variability was significantly lower for 3D-Slicer segmentations ($p = 0.0003$).

3D-Slicer segmentation process. To investigate the stability of 3D-Slicer algorithm against user seed-points initialization, we compared the intra-observer variability for each of the 3D-Slicer users. High overlap fractions were observed for the 3D-Slicer users: 95.01% ± 5.33%, 94.11% ± 3.95 and 97.08% ± 2.54% [mean ± SD], respectively.

To assess the duration of the 3D-Slicer segmentation process, we recorded the duration of all segmentation phases. The total segmentation times were in average 10.6 min (range: 4.85–18.25 min), 9.97 (range 6.39–13.83 min) and 9.94 min (range: 4.38–20.25 min), for the three 3D-Slicer users respectively. In average, the times measured for each 3D-Slicer segmentation phase were: loading (28 seconds), algorithm initialization (2.79 min), running the 3D-Slicer algorithm (32 seconds) and editing final phase (6.52 min).

Pathology. Further validation was provided by comparing the maximum diameter of the 3D slicer segmentations with that of the surgical specimen. Strong correlations were observed between the maximum diameter of 3D-Slicer volumes and the macroscopic diameter of the surgical tumours (spearman r, mean ± SD = 0.89 ± 0.05, range: 0.81–0.94). Similarly, the maximum diameters of the manual CT/PET delineations were highly correlated with the macroscopic diameter (spearman r, mean ± SD = 0.92 ± 0.02, range: 0.91–0.95). Figure 4 displays the scatter plot between macroscopic diameter and the diameters of CT segmentations (manual and 3D slicer). The diameters of surgery had a range of 1.8–9 and average of 4.5 ± 2.03 (mean ± SD). The manual delineations had a range of 1.42–12.53 and average of 6.09 ± 2.71 (mean ± SD). The semi-automatic delineations were: range 1.41–12.20 and average of 6.17 ± 2.89. These twelve different diameter

vectors were also compared using the Kruskal-Wallis test and no statistically significant difference was observed ($p = 0.97$).

Discussion

Despite the efforts in CT-PET imaging standardization and tumour delineation protocols, target definition remains subjected to observer variation. With respect to manual delineations, the addition of PET information to CT imaging in standardized delineation protocols has reduced the observer variability, however, human interaction and interpretation of medical images is still a considerable source of variation[3,22,23]. Furthermore, slice-by-slice manual contouring of two-dimensional images is a time consuming process.

Here, we evaluated the utility of a freely accessible 3D-Slicer algorithm, a cellular automaton-based algorithm, by performing a volumetric comparison with tumour delineations made by five independent oncologists following standardized protocols[24], as well as by comparing it with the maximal diameter obtained from pathological measurements.

The volumetric comparison showed that the 3D-Slicer algorithm provides tumour segmentations, statistically equivalent to physicians CT/PET manual contours. To evaluate the accuracy of the 3D-Slicer segmentations, the overlap fraction (%) was calculated and resulted in high values between the semi-automatically segmented volumes and the intersection (mean ± SD: 94.3 ± 4.4%, range: 76.8–99.8) and union (mean ± SD: 97.2 ± 5.1%; range: 72.6–100) of the manual delineations. Importantly, semi-automatic segmentations showed overall lower volume variability ($p = 0.0003$) and smaller uncertainty areas ($p = 0.0002$) compared to manual delineations. 3D-Slicer segmentations showed robustness towards user initialization, the OF's between the first Slicer segmentation and the second slicer segmentation were for each user in average: 95.01% ± 5.33%, 94.11% ± 3.95 and 97.08% ± 2.54%, respectively.

Additionally, we observed a strong correlation between the 3D-Slicer segmentations and the maximal diameter as measured on pathological examination (r = 0.89; 95% CI, 0.81–0.94). The average time to perform a complete segmentation was 9.8 minutes using Slicer. Loading the images and running the algorithm takes in average half a minute respectively. Due to the retrospective nature of our analysis we were not able to compare the 3D-Slicer segmentation times with the manual delineation times, since those were not available. However 3D slicer's volume segmentation has been shown to be substantially faster and less user intensive compared to manual delineation in other tumour sites[17]. Furthermore, manual delineation is well known to be a very time consuming task.

To minimize observer variability and reduce user interactions, several CT and PET semi-automatic segmentation methods have been introduced. Simple methods such as threshold-based segmentations are widely available but often fail to accurately define the tumour borders[10,11,16]. Various more complex methods have been investigated, including signal-to-background ratio individualized thresholding, watershed-based methods or complex fuzzy locally adaptive thresholding methods[11,14,15,25–27]. These methods have showed generally better correlations with pathology and manual delineations than the simple fixed threshold methods; however they often require significant tuning of algorithm parameters and are not widely available. PET-based methods are intrinsically better choices to segment the highly active metabolic areas of the tumour. In contrast, CT-based methods provide an anatomical segmentation with higher spatial resolution. In radiation therapy, CT is the reference imaging modality for treatment planning, and an accurate gross tumour volume definition is fundamental to assure adequate target coverage. Therefore, we believe that CT-based semi-automatic segmentations have clinical utility, if they provide segmentations as accurate as those generated manually by the medical experts, despite the intrinsic CT limitations to distinguish areas of the tumour that are metabolically more active.
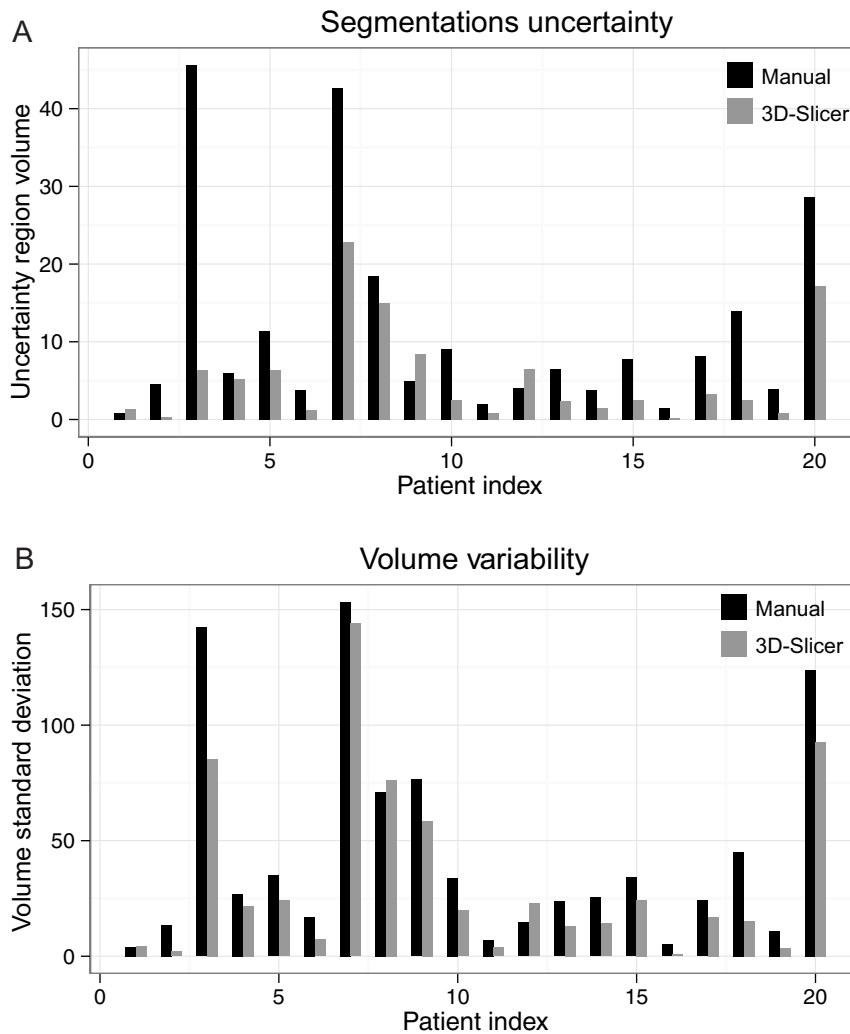
Figure 3 | (A): Comparison of volume uncertainty (as defined as the region that varied between the contours of multiple observers) of manual delineations and 3D-Slicer segmentations. See figure 2 for an illustrative example of the uncertainty region. (B): Comparison of volume variability (cm³) of observers' manual delineations and 3D-Slicer segmentations.

Cheebsumon et al, compared several commonly used PET-based segmentation methods with pathology and with a CT manually delineated volume[11]. They reported PET-based methods to have a better agreement with pathology compared to CT delineation. In their study, CT manual delineation significantly overestimated the tumour size compared to pathology. CT manual delineation is known to be prone to inter-observer variation and usually overestimates tumour dimensions. In their exhaustive methods comparison, they lacked a comparison with semi-automatic CT-based segmentation methods, which have shown better correlations with pathology than manual delineations[28]. We previously evaluated a CT-based click-and-grow ensemble segmentation (SCES) algorithm, which showed good overlap with medical expert's tumour delineations and with pathological measurements[28]. The SCES also showed robustness towards user initialization, as it involved an iterative segmentation process, with a bootstrapping routine with multiple initializations, which resulted in highly reproducible final segmentations[29]. Unfortunately, this algorithm is only available in commercial packages and therefore not available for the broader community.

A comparison of CT-based and PET-based methods with pathological measurements and manual delineations is still lacking though. We anticipate that methods combining CT and PET information will be the winner in the lung tumour segmentation race, though not all centers are equipped with integrated PET-CT scanners. However, intrinsic differences between CT and PET information should be taken into account. The present 3D-Slicer algorithm, provided accurate tumour segmentations for 85% of the cases. In three cases the 3D-Slicer failed to define accurately the border, these cases showed larger volume variability with 3D-Slicer compared to manual delineations; two of these cases were large masses with pleural attachment, however only one had a central location. The third case was a very small isolated tumour, adjacent to a main blood vessel, in this case due to the volume size, small variations in border definition due to the adjacent vessel, resulted in significant volume variations. Nevertheless, a medical expert should supervise auto-segmentation algorithms in all cases.

The current correlation between the 3D-Slicer delineation and pathology could possibly be improved if the CT and PET-CT would have been performed in 4D-mode. It is well recognized that a free-breathing CT and even more PET scan will result in blurred edges of the tumour and erroneous CT densities or SUV values. In further research, 4D scans should be used.

A general drawback when comparing segmentation algorithms with pathological dimensions is that often only tumour sizes in one dimension are available (maximal diameter). Furthermore, pathological measurements can be affected by tumour shrinkage and deformation after surgery. In this study only the maximal dia-
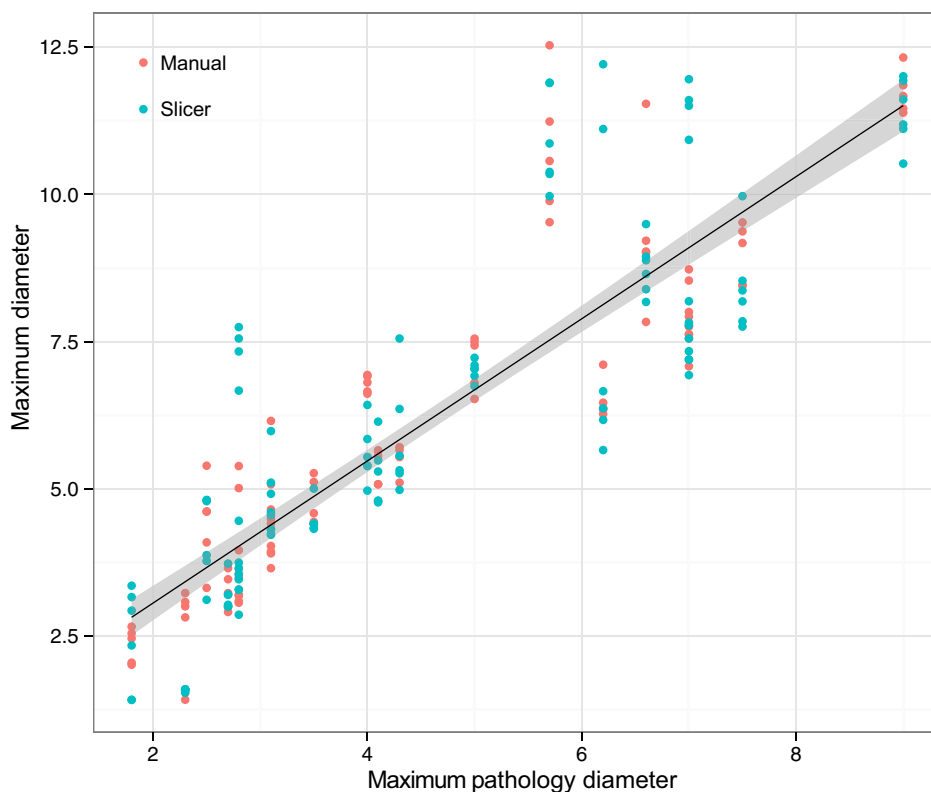
Figure 4 | Scatter plot between maximal diameter of surgical specimen and the maximal diameter of computed tomography (CT) segmented volumes for both manual and semiautomatic 3D-Slicer diameters. Spearman's correlation coefficient was 0.89 (95%CI, 0.81–0.94).

meter on pathology was compared, which is less prone to error than volumetric comparisons with pathology. The timing-span between the image acquisition and surgery may impact the comparison of the segmentation methods with pathology due to tumour growth. Given the correlation observed with pathological tumour diameter, this time difference may not have a strong impact in the evaluated cases.

In conclusion, the open source 3D-Slicer algorithm, provided tumour segmentations comparable to those manually delineated by physicians and with lower variability. Since the semi-automatic segmentations are statistically comparable to manual delineations and correlated well with pathology, they could be used as a starting point for treatment planning delineations and in high-throughput data mining research, such as Radiomics[18–21], where manual tumour delineations are often not available, or represent a considerable time consuming bottleneck.

Methods

CT-PET scans. The imaging data was acquired at MAASTRO Clinic in The Netherlands, as reported previously by Baardwijk et al[7]. In short, twenty consecutive patients with histologically verified non-small cell lung cancer, stage IB-IIIB, were included in this study. All patients received a diagnostic whole body positron emission tomography (PET)-computed tomography (CT) scanning (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. Free-breathing PET and CT images were acquired after a period of 45 minutes, during which the patient was encouraged to rest. The whole thorax spiral CT scan was acquired with intravenous contrast. The PET images were obtained 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. Imaging data are available on www.cancerdata.org. This study was conducted according to national laws and guidelines and approved by the appropriate local trial committee at Maastricht University Medical Center (MUMC+), Maastricht, The Netherlands. For more details see Baardwijk et al[7].

GrowCut semi-automatic segmentation method in 3D-Slicer. GrowCut is an interactive region growing segmentation method. Given an initial small set of label points the algorithm automatically segments the remaining image by using cellular automation. The algorithm uses a competitive region growing approach and is considered as having good accuracy and speed for the 2D and 3D image segmentation. For N-class segmentation the algorithm needs N initial sets of pixels (one set corresponding to each class) from user. Using these pixel sets, the algorithm automatically generates the region of interest (ROI), which is the convex hull of the user-labelled pixels with an additional margin. In the next step, it iteratively labels all the pixels in the ROI using the user-given pixel labels. The algorithm converges when all the pixels in the ROI have unchanged labels across several iterations. Pixel labelling is done using a weighted similarity score, which is a function of the neighbouring pixel
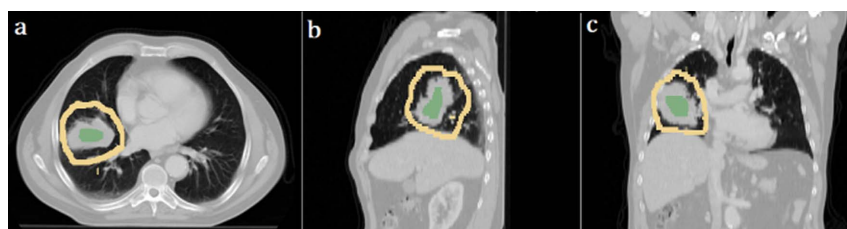

Figure 5 | Initialization step of 3D-Slicer segmentation. Marked foreground (green) and background (yellow) are shown. Axial (a), sagittal (b) and coronal (c) views are shown.
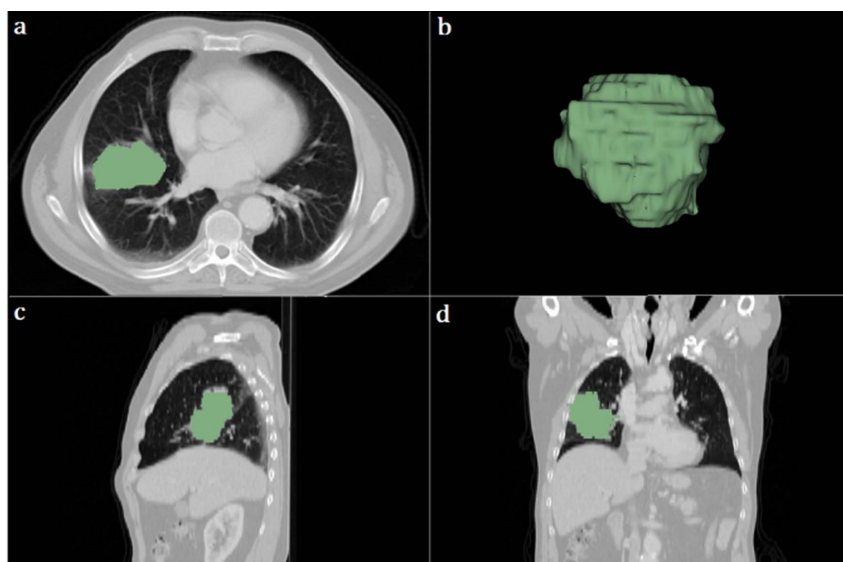
Figure 6 | Semi-automatically segmented tumour (green) using 3D-Slicer. Axial (a), three dimensional (b), sagittal (c) and coronal (d) views are shown.

weights. An unlabelled pixel is labelled corresponding to the neighbouring pixels that have the highest weights.

NSCLC tumour GrowCut segmentation in 3D-Slicer. 3D-Slicer gives a user friendly GUI as the frontend and an efficient algorithm as the backend for the GrowCut segmentation. After loading the patient data, the process began with the initialization of the foreground and background by marking the area inside and outside the tumour region with few initial seed pixels [Figure 5]. The next step was automatic competing region-growing, which segmented the region of interest into foreground and background. Background and surrounding isolated foreground pixels were removed after visual inspection. Figure 6 displays the final segmented tumour region. In Supplementary Figure S2 four representative tumour segmentations generated using the 3D-Slicer algorithm are compared with the manual delineations of five independent observers. Visual comparison shows a high agreement of the manual delineations with the semiautomatic one.

We performed Slicer GrowCut segmentations by three independent users, which repeated the process two times, with a three day interval between each time. Segmentation times using GrowCut were recorded for every step of the analysis.

Manual tumour delineations. To validate the semiautomatic segmentation method, five radiation oncologist have manually delineated the gross tumour volume (GTV) of the primary tumour, based on fused PET-CT images using standard delineation protocol, which includes 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)[2,7,24]. Radiation oncologists were mutually blind of each other's delineations. The primary GTV was defined for each patient based on combined CT and PET information in the axial plane. The radiation oncologists were given transversal, coronal, sagittal and 3D views simultaneously. A treatment planning system (XiO; Computer Medical System, Inc., St. Louis, MO), was used for performing delineations.

Pathology. The examination of surgical specimen was carried out according to national guidelines[7]. Surgical resections were performed on all the patients. Before slicing, the maximal diameter of the primary tumour 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).

Statistical analysis. Overlap Fraction (OF) was used to evaluate the 3D slicer's segmentations in terms of its spatial overlap with manual delineations. Intersection and union volumes were defined for manual delineations (Figure 1). OFs were calculated between the semiautomatic segmentations and these intersection and union delineations. OF was defined as the as the volume of overlap divided by the smallest volume[30]:

$$OFinter = \frac{SV \cap OBi}{\min\{SV, OBi\}} * 100, \qquad and$$

$$OFunion = \frac{SV \cap OBu}{\min\{SV, OBu\}} * 100$$

SV, OBi and *OBu* are the semiautomatic, observers' intersection and union volumes respectively. OF value 100 suggests perfect match while OF value 0 points to two disjoint volumes and thus no match. OF_{inter} indicates whether the semiautomatic-segmentation

method covers the common agreement (intersection volume) of the manual delineations while OF_{union} indicates whether the algorithm falls within the inter-observer variability (union volume).

Furthermore, using the above described concept of union and intersection volumes, we calculated and compared the uncertainty of the GrowCut segmentations and the manual delineations. The uncertainty was defined as the difference between the union and intersection volumes, which is the area that belongs to the union but not to the intersection volumes. This region can be seen in Figure 1, highlighted in green. The lower the difference between union and intersection volumes the lower the uncertainty. If all contours were equal, with no variation, the union and intersection volumes would be identical with no uncertainty areas.

Overlap fractions were used to compare the first 3D-Slicer segmentation against the second 3D-Slicer segmentation for the same observer.

A volume (cm³) comparison was also carried out. Volumes calculated from different segmentation methods were compared using the Kruskal-Wallis test. Two methods were considered to be significantly different when the p-value was lower than 0.05.

We compared the volume variability of the 3D-Slicer segmentations against manual delineations using the standard deviation of the 3D-Slicer and manual volumes. The Wilcoxon test was used to compare the volume variability and uncertainty differences between the two types of segmentations.

Spearman correlation coefficient was used to compare the maximal diameter of pathology with the maximal diameter of 3D-Slicer and the manual segmentations. Further we also compared all these twelve maximal diameter groups: 3D-Slicer (three observers twice), pathology, and five manual using the Kruskal-Wallis one-way analysis of variance. Again groups were considered significantly different when the p-value was lower than 0.05. All data are expressed as mean ± SD. All the analyses were performed in Matlab (The MathWorks Inc., Natick, MA, USA) and R (R Foundation for Statistical Computing, Vienna, Austria).

1. Jemal, A. *et al.* Global cancer statistics. *CA Cancer J Clin* **61**, 69–90 (2011).
2. van Baardwijk, A. *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* **28**, 1380–1386 (2010).
3. Steenbakkers, R. J. *et al.* Observer variation in target volume delineation of lung cancer related to radiation oncologist-computer interaction: a 'Big Brother' evaluation. *Radiother Oncol* **77**, 182–190 (2005).
4. Van de Steene, J. *et al.* Definition of gross tumor volume in lung cancer: inter-observer variability. *Radiother Oncol* **62**, 37–49 (2002).
5. Bowden, P. *et al.* Measurement of lung tumor volumes using three-dimensional computer planning software. *Int J Radiat Oncol Biol Phys* **53**, 566–573 (2002).
6. Caldwell, C. B. *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* **51**, 923–931 (2001).
7. van Baardwijk, A. *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* **68**, 771–778 (2007).
8. Steenbakkers, R. J. *et al.* Reduction of observer variation using matched CT-PET for lung cancer delineation: a three-dimensional analysis. *Int J Radiat Oncol Biol Phys* **64**, 435–448 (2006).

9. De Ruysscher, D. PET-CT in radiotherapy for lung cancer. *Methods Mol Biol* **727**, 53–58 (2011).

10. Nestle, U. *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* **46**, 1342–1348 (2005).

11. Cheebsumon, P. *et al.* Assessment of tumour size in PET/CT lung cancer studies: PET- and CT-based methods compared to pathology. *EJNMMI Res* **2**, 56 (2012).

12. Daisne, J. F. *et al.* Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. *Radiother Oncol* **69**, 247–250 (2003).

13. Dehmeshki, J., Amin, H., Valdivieso, M. & Ye, X. Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach. *IEEE Trans Med Imaging* **27**, 467–480 (2008).

14. Schaefer, A. *et al.* PET-based delineation of tumour volumes in lung cancer: comparison with pathological findings. *Eur J Nucl Med Mol Imaging* **40**, 1233–1244 (2013).

15. Hatt, M. *et al.* Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. *Int J Radiat Oncol Biol Phys* **77**, 301–308 (2010).

16. Wu, K. *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* **77**, 699–706 (2009).

17. Egger, J. *et al.* GBM volumetry using the 3D Slicer medical image computing platform. *Sci Rep* **3**, 1–7 (2013).

18. Kumar, V. *et al.* Radiomics: the process and the challenges. *Magn Reson Imaging* **30**, 1234–1248 (2012).

19. Lambin, P. *et al.* Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* **48**, 441–446 (2012).

20. Buckler, A. J., Bresolin, L., Dunnick, N. R. & Sullivan, D. C. A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging. *Radiology* **258**, 906–914 (2011).

21. Buckler, A. J. *et al.* Quantitative imaging test approval and biomarker qualification: interrelated but distinct activities. *Radiology* **259**, 875–884 (2011).

22. Greco, C., Rosenzweig, K., Cascini, G. L. & Tamburrini, O. Current status of PET/CT for tumour volume definition in radiotherapy treatment planning for non-small cell lung cancer (NSCLC). *Lung Cancer* **57**, 125–134 (2007).

23. Sonke, J. J. & Belderbos, J. Adaptive radiotherapy for lung cancer. *Semin Radiat Oncol* **20**, 94–106 (2010).

24. van Baardwijk, A. *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* **71**, 1394–1401 (2008).

25. Ye, X., Beddoe, G. & Slabaugh, G. Automatic Graph Cut Segmentation of Lesions in CT Using Mean Shift Superpixels. *Int J Biomed Imaging* **2010**, 1–14 (2010).

26. Daisne, J.-F. o. *et al.* Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. *Radiotherapy and Oncology* **69**, 247–250 (2003).

27. Wanet, M. *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* **98**, 117–125 (2010).

28. Rios Velazquez, E. *et al.* A semiautomatic CT-based ensemble segmentation of lung tumors: comparison with oncologists' delineations and with the surgical specimen. *Radiother Oncol* **105**, 167–173 (2012).

29. Gu, Y. *et al.* Automated Delineation of Lung Tumors from CT Images Using a Single Click Ensemble Segmentation Approach. *Pattern Recognit* **46**, 692–702 (2013).

30. Aerts, H. J. *et al.* Identification of residual metabolic-active areas within individual NSCLC tumours using a pre-radiotherapy (18)Fluorodeoxyglucose-PET-CT scan. *Radiother Oncol* **91**, 386–392 (2009).

Acknowledgments

Author contributions

H.J.W.L.A., E.R.V. and C.P. conceived of the project, analysed the data, and wrote the paper. M.J., R.H.M., A.B., F.M.F., J.H.L., D.D.R., R.K. and P.L. provided expert guidance, data, or analysis tools and reviewed the manuscript.

Additional information