Lung Cancer 82 (2013) 76–82

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan

Inter-observer reproducibility of semi-automatic tumor diameter measurement and volumetric analysis in patients with lung cancer

J. Dinkel[a,b,c,*], O. Khalilzadeh[d], C. Hintze[e], M. Fabel[e], M. Puderbach[c,f], M. Eichinger[b,c], H.-P. Schlemmer[b], M. Thorn[g], C.P. Heussel[c,f], M. Thomas[h], H.-U. Kauczor[a,c], J. Biederer[a,c,e]

[a] Department of Radiology, University Hospital Heidelberg, Heidelberg, Germany
[b] Department of Radiology, German Cancer Research Center, Heidelberg, Germany
[c] Translational Lung Research Center Heidelberg, Member of the German Center for Lung Research, Heidelberg, Germany
[d] Department of Radiology, Massachusetts General Hospital, Boston, MA, USA
[e] Department of Diagnostic Radiology, University Hospital Schleswig-Holstein, Kiel, Germany
[f] Department of Radiology, Thoraxklinik Heidelberg, University Hospital Heidelberg, Heidelberg, Germany
[g] Siemens Oncology, Forchheim, Germany
[h] Department of Oncology, Thoraxklinik Heidelberg, University Hospital Heidelberg, Heidelberg, Germany

ARTICLE INFO

Article history:
Received 21 March 2013
Received in revised form 20 June 2013
Accepted 7 July 2013

Keywords:
Lung cancer
Volumetric analysis
Inter-rater reliability
Tumor size
WHO guideline
RECIST guideline

ABSTRACT

Objectives: Therapy monitoring in oncologic patient requires precise measurement methods. In order to improve the precision of measurements, we used a semi-automated generic segmentation algorithm to measure the size of large lung cancer tumors. The reproducibility of computer-assisted measurements were assessed and compared with manual measurements.
Methods: CT scans of 24 consecutive lung cancer patients who were referred to our hospital over a period of 6 months were analyzed. The tumor sizes were measured manually by 3 independent radiologists, according to World Health Organization (WHO) and the Revised Response Evaluation Criteria in Solid Tumors (RECIST) guidelines. At least 10 months later, measurements were repeated semi-automatically on the same scans by the same radiologists. The inter-observer reproducibility of all measurements was assessed and compared between manual and semi-automated measurements.
Results: Manual measurements of the tumor longest diameter were significantly ($p < 0.05$) smaller compared with the semi-automated measurements. The intra-rater correlations coefficients were significantly higher for measurements of longest diameter (intra-class correlation coefficients: 0.998 vs. 0.986; $p < 0.001$) and area (0.995 vs. 0.988; $p = 0.032$) using semi-automated compared with manual method. The variation coefficient for manual measurement of the tumor area (WHO guideline, 15.7% vs. 7.3%) and the longest diameter (RECIST guideline, 7.7% vs. 2.7%) was 2–3 times that of semi-automated measurement.
Conclusions: By using computer-assisted size assessment in primary lung tumor, interobserver-variability can be reduced to about half to one-third compared to standard manual measurements. This indicates a high potential value for therapy monitoring in lung cancer patients.

1. Introduction

Precision and reproducibility of tumor size measurements is a critical point for planning and monitoring oncologic therapy. By contemporary standards, the response to treatment is monitored with CT or MRI using robust criteria, since small differences in the assessment of therapeutic response may influence the overall outcome of clinical trials. The criteria of the World Health Organization (WHO) presented 30 years ago are defined as the product between the maximum axial tumor diameter and its longest perpendicular diameter in the same image [1]. The RECIST criteria (Revised Response Evaluation Criteria in Solid Tumors), introduced in the 1990s, rely mainly on the sum of unidimensional measurements of the longest axial diameter [2,3]. Numerous studies demonstrated that RECIST criteria are a reliable and simple measurement for quantifying therapy response [4,5]. The RECIST criteria were lately revised to address some issues regarding number of target lesions and measurement of lymph nodes [6,7].

Since the TNM staging system relies partly on the size of the primary tumor, choice of treatment and prognosis are associated with

* Corresponding author at: Department of Radiology, University Hospital Heidelberg, INF 110, 69120 Heidelberg, Germany. Tel.: +49 6221 56 35 169.
E-mail addresses: julien.dinkel@med.uni-heidelberg.de, julien.dinkel@gmail.com (J. Dinkel).

0169-5002/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.lungcan.2013.07.006

the extend of the tumor [8]. Furthermore, an accurate measurement is of utmost importance for the follow-up and therapy monitoring. However, the precision of tumor measurement is influenced by imaging technology as well as by the subjects of the study. Even if the examination parameters are kept as similar as possible (including CT reconstruction kernel, windowing and slice thickness), slight variations in slice position or different patient orientation, may considerably influence the results of the measurements [9]. In addition, the variability introduced by individual examiners may lead to divergent classifications of tumor response [10] with potential therapeutic consequences [11]. Therefore, methods for precise and reproducible tumor measurements are of utmost clinical importance [12].

For the analysis of small lung nodules, studies have demonstrated that computer aided methods improve accuracy [13,14], reproducibility [15,16] and decreased interobserver variability [15,17] of volumetric assessment. Lung nodule automated segmentation algorithms are optimized for screening purposes and usually work for small nodules (with a diameter smaller than 2 cm). The automated analysis of larger lung tumors is, however, much more challenging due to their usual irregular shape. In addition, they are often located close to structures of similar density (i.e. mediastinum, pleura).

Considering the high clinical significance of lung cancer therapy monitoring, precise methods for measurement of these tumors are particularly important [12]. In this study, we used a generic segmentation algorithm that works for unspecific tissues and/or large tumors. We assumed that this non-specific segmentation algorithm would be very useful for the size assessment of large lung cancer tumors. It was therefore the purpose of our study to compare the reproducibility of computer assisted uni- and bi-dimensional measurements with manual measurements based on RECIST and WHO criteria for the assessment of the primary lung cancer tumors. As alternative measurement with potential value for future study designs, volumetric quantification of the lung tumors was also tested and its reproducibility was assessed.

2. Patients and methods

2.1. Patients' characteristics

This study was approved by our institutional review board. The chest CT datasets of 24 consecutive patients, who were admitted to our hospital between 3.2010 and 9.2010 and had a biopsy-proven lung cancer of 1 cm or larger, were selected for this study. Patients have been referred for staging prior to non-surgical oncological treatment. Patients with irregularly shaped tumors, with contact to pleura and vessels were also included.

2.2. CT data acquisition

A multi-slice CT scan was performed using a standard dose technique for clinical purposes. All scans were acquired with a 4-slice multi-slice CT scanner (Somatom Sensation 4, Siemens Medical, Forchheim, Germany) using spiral mode scanning of the entire chest in the craniocaudal direction. The following scan parameters were used for patients in supine position at full inspiration: 210 mAs (range 110–270 mAs) tube current with a pitch of 1.5 at a tube voltage of 120 kV, 45 s after injection of 80 ml contrast material (3 ml/s; Imeron 300; Bracco, Milan, Italy). Four patients with contra-indication to iodinated contrast medium underwent a non-contrast CT-scan. Examinations were reconstructed with 2.5 mm section thickness, 2.5 mm reconstruction interval using a medium soft tissue kernel (B30) a typical field of view of 380 mm and a 512×512 matrix.

2.3. Tumor measurement

Pulmonary tumors were analyzed independently by 3 radiologists (with 4, 5 and 7 years of experience in radiology) on axial reconstructions. The images were displayed on a workstation (Leonardo workstation, Siemens Medical Solutions, Forchheim, Germany) using a lung window with a width and center setting of 1500 and −500 HU, respectively. The readers were allowed to change window settings, if necessary. The readers performed manual bi-dimensional measurements on transverse slices using a digital caliper according to the RECIST and WHO criteria: the longest diameter; the longest perpendicular diameter in the same image; and the product of these two diameters ("WHO area").

Then, the readers performed a computer aided semi-automated evaluation of the tumors using dedicated software: Syngo Oncology (Siemens Medical Solutions, Forchheim, Germany) at least 10 months after the manual readings in order to avoid recall bias. Four measurements were semi-automatically generated: the longest diameter on native axial slice; the longest perpendicular diameter in the same image; the product of these two diameters ("WHO area"); and the volumetric quantification of the tumor. The observers were not aware of each other's selected slices.

The software used the segmentation method described elsewhere [16,19]. We used the generic non-specific segmentation algorithm instead of the lung nodule algorithm. The lung nodule algorithm is optimized for small structures (<2 cm) and was therefore not applicable in our study. The procedure accounts for density variations depending on contrast agent timing, cancer type, surrounding atelectasis or thorax wall. The user draws a rough diameter across the lesion in one image plane. The algorithm estimates thresholds by histogram analysis around the lesion. Then, an initial segmentation is generated using both 3D region-growing technique and estimated thresholds. Adjacent structures of similar density are separated by morphological operations. A plausibility check is used to validate the correctness of the final segmentation based on the initial diameter. Eventually, the reader is able to verify the segmentation quality in three orthogonal views of the lesion and in a 3D volume reconstruction of the segmentation result (Fig. 1). The accuracy of the segmentation was not evaluated, since the readers were allowed to adjust the segmentation if necessary, in order to achieve the most satisfactory results. In case of mis-segmentation, the initial diameter was repositioned and/or, if necessary, the segmented volume was semi-automatically modified to match the visual assessment of the tumor.

2.4. Statistical analysis

Statistical analysis was performed with R, version 2.13 (R Foundation for Statistical Computing, Vienna, Austria) and GraphPad Prism (version 6.00 for Mac, GraphPad Software, La Jolla, California, USA). The required sample size to detect a significant association at $\alpha = 0.05$ and with a power of 80% was estimated to be 24. Continuous variable are expressed as mean ± standard deviation (SD), or median and interquartile range. Paired sample *T*-test was used to compare the semi-automated and manual measurements. Intra-class correlation coefficients between measurements of different raters and between manual and semi-automated methods were calculated. Intraclass-correlation coefficients were compared between manual and semi-automated methods using the method described by Feldt et al. [20].

Linear regression method was also used to fit regression lines and evaluate the correlation between the measurements of the raters using manual and semi-automated methods. The variation coefficient (VC), defined as the ratio of the standard deviation to the mean, was calculated as an index for the reproducibility of manual and semi-automated tumor measurement, including 3D volumes,