

Effect of Kernels Used for the Reconstruction of MDCT Datasets on the Semi-Automated Segmentation and Volumetry of Liver Lesions

Auswirkungen der zur Rekonstruktion von CT-Datensätzen verwendeten Kernel auf halbautomatische Segmentierung und Volumetrie von Leberläsionen

Authors

D. Pinto dos Santos¹, R. Kloeckner¹, K. Wunder¹, L. Bornemann², C. Düber¹, P. Mildenberger¹

Affiliations

¹ Department of Diagnostic and Interventional Radiology, University Medical Center of the Johannes Gutenberg University Mainz

² MEVIS, Fraunhofer, Bremen

Key words

- CT
- image manipulation/reconstruction
- technical aspects

Zusammenfassung



Ziel: Untersuchung der Auswirkungen verschiedener Rekonstruktionkernel auf die halbautomatische Segmentierung von Leberläsionen in der MDCT.

Material und Methoden: Insgesamt 62 Leberläsionen wurden durch drei unabhängige Radiologen mithilfe der Oncology-Prototype Segmentierungssoftware (Fraunhofer MEVIS, Siemens Healthcare, Germany) halbautomatisch vermessen. Die verwendeten CT-Datensätze waren jeweils mittels standard, soft und detailed (Philips B, A and D) Kernel rekonstruiert worden. Um eine objektive Messung sicherzustellen wurden nur Läsionen eingeschlossen, deren initiale Segmentierung zufriedenstellend war, manuelle Korrekturen wurden nicht vorgenommen. Effektiver Durchmesser und Volumen wurde für alle Läsionen erhoben. Die Segmentierung in den mittels soft und detailed Kernel rekonstruierten Datensätzen wurde durch Kopieren der Seedposition aus dem Standardkernel Datensatz vorgenommen.

Ergebnisse: Der mittlere effektive Läsionsdurchmesser betrug bei Verwendung des Standardkernels $19,9 \pm 9,7$ mm. Der Vergleich aller drei Kernel untereinander zeigte keine signifikanten Unterschiede. Der mittlere Unterschied zwischen standard und soft Kernel betrug $1\% \pm 6\%$, zwischen standard und detailed $3\% \pm 13\%$ und zwischen soft und detailed $2\% \pm 9\%$. Die Intra-Klassen-Korrelationskoeffizienten lagen für alle Vergleiche bei $> 0,96$.

Schlussfolgerung: Die halbautomatische Segmentierung und Volumetrie von Leberläsionen zeigt verlässliche Messungen unabhängig von dem zur Rekonstruktion der MDCT-Datensätze verwendeten Kernel.

Kernaussagen:

- ▶ Die halbautomatische Segmentierung und Volumetrie von Leberläsionen ist verlässlich unab-

Abstract



Purpose: To evaluate the effect of different reconstruction kernels on the semi-automated segmentation of liver lesions in MDCT.

Materials and Methods: A total 62 liver lesions were measured by three independent radiologists with the semi-automated segmentation software Oncology-Prototype (Fraunhofer MEVIS, Siemens Healthcare, Germany) using MDCT datasets (3-mm slice thickness, 2-mm increment) reconstructed with standard, soft and detailed kernels (Philips B, A and D). To ensure objective measurements, only lesions with satisfactory initial segmentation were included, and manual correction was not used. The effective diameter and volume were recorded for each lesion. Segmentation in the soft and detailed kernel datasets was performed by copying the initial seed's position from the standard kernel dataset.

Results: The mean effective lesion diameter was 19.9 ± 9.7 mm using the standard kernel. Comparing the three kernels, no significant differences were found. The mean difference was $1\% \pm 6\%$ for the standard kernel compared to the soft kernel, $3\% \pm 13\%$ for the standard kernel vs. the detailed kernel and $2\% \pm 9\%$ for the soft kernel compared to the detailed kernel. The intra-class correlation coefficients were > 0.96 in all cases.

Conclusion: The semi-automated segmentation and volumetry of liver lesions shows reliable measurements regardless of the kernel used for reconstruction of the MDCT dataset.

Key Points:

- ▶ Semi-automated segmentation and volumetry of liver lesions is reliable regardless of the kernel used for reconstruction of the MDCT dataset.
- ▶ Until today the gold standard for the evaluation of tumor response has been unidimensional manual measurement.
- ▶ Volumetric measurements could improve the assessment of tumor growth.

received 7.1.2013
accepted 23.10.2013

Bibliography

DOI <http://dx.doi.org/10.1055/s-0033-1356178>
Published online: 23.1.2014
Fortschr Röntgenstr 2014; 186: 780–784 © Georg Thieme Verlag KG Stuttgart · New York · ISSN 1438-9029

Correspondence

Dr. Daniel Pinto dos Santos
Department of Diagnostic and Interventional Radiology,
University Medical Center of the Johannes Gutenberg University Mainz
Langenbeckstraße 1
55131 Mainz
Germany
Tel.: ++49/61 31/17 60 11
Fax: ++46/61 31/64 43
pintodos@uni-mainz.de

hängig von dem zur Rekonstruktion der MDCT-Datensätze verwandten Kernel.

- Bis heute ist der Gold-Standard zur Evaluation von Tumoransprechen die manuelle unidimensionale Messung.
- Volumetrische Messungen könnten die Beurteilung von Tumorstadium verbessern.

Introduction

Evaluation of tumor response is crucial for oncological trials and adequate therapy management in oncological patients. In solid tumors this is mainly based on assessing changes in lesion size with the Response Evaluation Criteria in Solid Tumors (RECIST) in the original and modified version (mRECIST) being a widely accepted, reliable and reproducible way of doing so [1, 2]. Targeted lesions in multidetector computed tomography (MDCT) are measured either by taking the longest diameter in the axial plane for extranodal lesions or the short axis for lymph nodes. Disease status is then assessed as “progressive”, “stable”, “partial” or “complete response” based on the sum of the diameters compared to the preceding studies. However, manual measurements are subject to inter- and intraobserver variability and moreover anisotropic changes in tumor size might be missed [3, 4]. To improve the reproducibility and accuracy of measurements as well as assessing tumor volume changes, software applications offering semi-automated lesion segmentation have been developed [5]. In recent studies these tools showed promising results with respect to segmentation and volumetry of lymph nodes [6–8], lung nodules [9] and liver lesions [10–13].

In the clinical routine and even more so in retrospective oncological trials, there can be significant differences in the datasets used for semi-automated segmentation. Varying CT scanning parameters and slice thickness as well as the kernel used for reconstruction of datasets might have impact on the accuracy of lesion measurement. For in vivo conditions Puesken et al. showed that slice thickness and CT dose levels had some influence on semi-automated measurements with deterioration of results in slice thicknesses greater than 3 mm [14]. In phantom liver lesions low variability of measurements was shown with varying CT scanning parameters [15]. However, no data is available on kernel-dependent variability in vivo.

Therefore, the purpose of this study was to investigate the influence of different reconstruction kernels on semi-automated segmentation and volumetry using the “Oncology Prototype” software (Fraunhofer MEVIS, Siemens Healthcare, Germany).

Materials and Methods



Patients

A total of 27 patients (10 men, 17 women; median age: 60.2 years, min/max: 30.8/81.8 years) with histopathologically confirmed metastatic tumor disease were recruited for this study from June to August 2010. The tumor entities included colorectal carcinoma (n=7), lung cancer (n=7), breast cancer (n=3) and others (n=10). All patients were referred to our department as indicated by clinical needs and received MDCT for follow-up or further investigation. Therefore, approval by the institutional review board was not necessary. The inclusion criteria were: solitary or multifocal hypodense liver lesions in portal venous phase of contrast-enhanced MDCT, with at least one well delimited le-

Citation Format:

- Pinto dos Santos D, Klöckner R, Wunder K et al. Effect of Kernels Used for the Reconstruction of MDCT Datasets on the Semi-Automated Segmentation and Volumetry of Liver Lesions. *Fortschr Röntgenstr* 2014; 186: 780–784

sion. The exclusion criteria were non-diagnostic CT due to significant motion artifacts or inappropriate bolus timing hampering the appropriate evaluation.

Imaging parameters

All MDCT imaging was performed with either a 256-row or 64-row CT scanner (iCT 256® or Brilliance 64®, Philips Medical Systems, The Netherlands). The tube voltage was 120 kV with 150 mAs and longitudinal dose modulation in both cases (mean CTDI/DLP was 9.8 mGy/712.9 mGy*cm for iCT 256® and 9.6 mGy/532.3 mGy*cm for Brilliance 64®). Scans were performed during breath-hold with a fixed delay of 60 s (portal venous phase) after intravenous administration of 100 ml of nonionic iodine contrast agent at a flow rate of 3 ml/s (Imeron 300, Bracco Imaging, Milan, Italy) via a power injector (Injectron CT2, Medtronic AG, Germany). Datasets were acquired with 64 × 0.625-mm collimation and a rotation time of 0.75 s in both cases. The pitch was 1.172 for the 64-row scanner and 0.984 for the 256-row scanner. During the study period all CT scans of patients with confirmed metastatic disease and one or multiple hypodense liver lesions were prospectively reconstructed with a 3-mm slice thickness, a 2-mm increment using soft, standard and detailed reconstruction kernels (Philips A, B, and D) and a soft-tissue window (center/width 60/360 HU).

Liver lesions

The reconstructed CT datasets were transferred to a separate workstation for segmentation with the “Oncology Prototype” software (Fraunhofer MEVIS, Siemens Healthcare, Germany). To assure objective evaluation of measurement differences depending on the reconstruction kernels used, manual correction of the software’s segmentation was not allowed. Therefore, an unblinded radiologist selected a total of 62 well delimited hypodense liver lesions with a minimum diameter of 5 mm and a satisfactory initial segmentation result (► Fig. 1). These lesions were digitally tagged for further evaluation by the readers. A mean of 2.82 lesions were included per patient. All lesions were numbered to avoid correlation/mapping errors between the different readers and datasets.

Semi-automated segmentation

All 62 previously tagged lesions were independently evaluated by three blinded readers (experienced in computed tomography and oncological imaging) in randomized sequence using the “Oncology Prototype” software (Fraunhofer MEVIS, Siemens Healthcare, Germany). The software uses a dedicated algorithm for the segmentation of liver lesions based on lesion density, image noise and contrast difference with respect to the adjacent liver parenchyma. After an initial diameter is drawn by the reader in an arbitrary slice on a liver lesion, region-growing-based algorithms are used in combination with watershed transformation and distance transformation algorithms for the segmentation and separation of adjacent structures of similar density (e.g. vessels, bile ducts or adjacent adipose tissue) [16]. As the purpose of this study was to analyze the influence of different reconstruc-

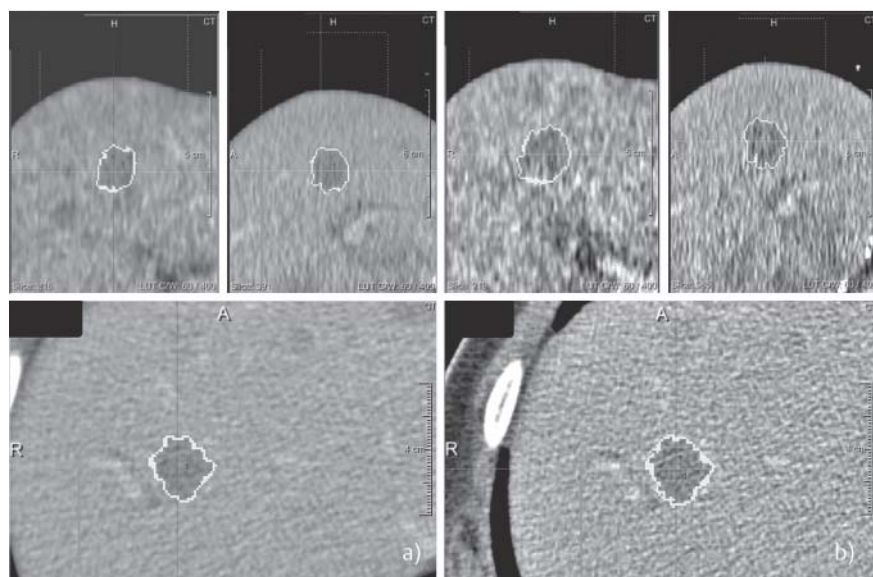


Fig. 1 Segmentation result as shown by the software **a)** B kernel (standard) and **b)** D kernel (detailed). The lesion is shown in coronal and sagittal view (top left corner and top right corner) as well as in the axial plane (bottom viewport). The lesion edges are highlighted in the different viewports.

Abb. 1 Segmentierungsergebnis im Viewer der Software **a)** kernel B (standard) und **b)** kernel D (detailed). Die Läsion wird sowohl koronar und sagittal (obere linke und obere rechte Ecken) sowie in der axialen Schicht (unterer Viewport) dargestellt. Die Läsionsränder sind im jeweiligen Viewport hervorgehoben.

tion kernels on semi-automated segmentation, any reader-dependent influence had to be excluded. Therefore, the first segmentation was performed on the standard kernel dataset. Subsequently, the exact seed diameter position was copied to the other kernel datasets using the software's dedicated tool. The volume, effective diameter and auto-RECIST were recorded for further analysis. The measured volume corresponds to the number of voxels included by the segmentation algorithm multiplied by voxel size. The effective diameter and auto-RECIST are derived values, where the effective diameter is defined as $eD = (\frac{6 \cdot V}{\pi})^{\frac{1}{3}}$ and auto-RECIST is computed by pairwise comparison of outer border voxels in any one given axial plane.

Data analysis

To avoid bias, statistical analysis was performed by an independent statistician at the hospital's Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI) using the software SAS® (SAS Institute Inc., Chicago, USA) and SPSS® (Version 17 for Windows, SPSS Inc., Chicago, USA). All analyses were intended to be exploratory, and p-values < 0.05 were regarded as significant. Results were compared using standard descriptive statistical analysis and the Wilcoxon signed-rank test for paired data. Concordance between measurements in different kernel datasets as well as interobserver variations were analyzed using intraclass correlation coefficients (ICC) and Bland-Altman plots [17, 18]. Since no information on the true size/volume of the lesions was available, the mean of the measurements in the standard kernel was used as a reference.

Results

Liver lesions

All 62 lesions of the 22 included patients (mean 2.82 ± 1.79 lesions per patient) had satisfactory initial segmentation results in the standard kernel for all three readers. The lesion's effective diameter measured using the standard kernel ranged from 5.6 mm to 51.7 mm (median 18.2 mm). The lesion's mean density values

ranged from 4.99 HU to 97.08 HU (median 54.52 HU). All lesions could successfully be segmented in the standard kernel. A total of 12 lesions showed poor segmentation results after exactly copying the seed diameter to a different kernel using the software's dedicated tool in some cases. These cases seemed randomly distributed over the different kernels and readers and no reproducible cause could be identified. However, these lesions could be successfully segmented when drawing a new initial seed diameter in the respective kernel. Nevertheless, to prevent bias, these measurements were excluded from further analysis.

Semi-automated segmentation

No significant differences were found between the measurements in the different kernel datasets. The median effective diameter per lesion in the standard kernel was 18.3 mm (min/max 6.2 mm/49.9 mm), and the mean difference in the soft kernel was $1.0 \pm 4.8\%$ (median 18.2 mm, min/max 6.7 mm/50.6 mm, $p=0.198$) and $1.4 \pm 7.6\%$ (median 18.7 mm, min/max 6.9 mm/50.9 mm, $p=0.148$) in the detailed kernel. Similar results were found for lesion volumes, where the relative difference seemed bigger as it is the third power of a radius. The median lesion volume in the standard kernel was 3.2 ml (min/max 0.1 ml/65.8 ml), and the mean difference in the soft kernel was $3.66 \pm 15.65\%$ (median 3.2 ml, min/max 0.1 ml/67.9 ml, $p=0.379$) and $5.41 \pm 25.32\%$ (median 3.5 ml, min/max 0.2 ml/69.1 ml, $p=0.235$) in the detailed kernel. Interestingly, the differences were smaller when evaluating by auto-RECIST measurements. The median auto-RECIST in the standard kernel was 20.6 mm (min/max 5.9 mm/73.2 mm). Compared to the standard kernel, the mean difference for auto-RECIST in the soft kernel was $0.49 \pm 6.68\%$ (median 20.4 mm, min/max 6.5 mm/68.5 mm, $p=0.798$) and $0.71 \pm 7.66\%$ (median 20.6 mm, min/max 4.7 mm/68.3 mm, $p=0.651$) in the detailed kernel.

All measurements were highly correlated. We compared every possible combination of observers and kernels, e.g. observer 1 kernel B with observer 2 kernel B, but also observer 1 kernel D with observer 3 kernel A. The intraclass correlation coefficients (ICCs) ranged from 0.96 to 0.99 for all comparisons between kernels and/or observers (► Fig. 2).

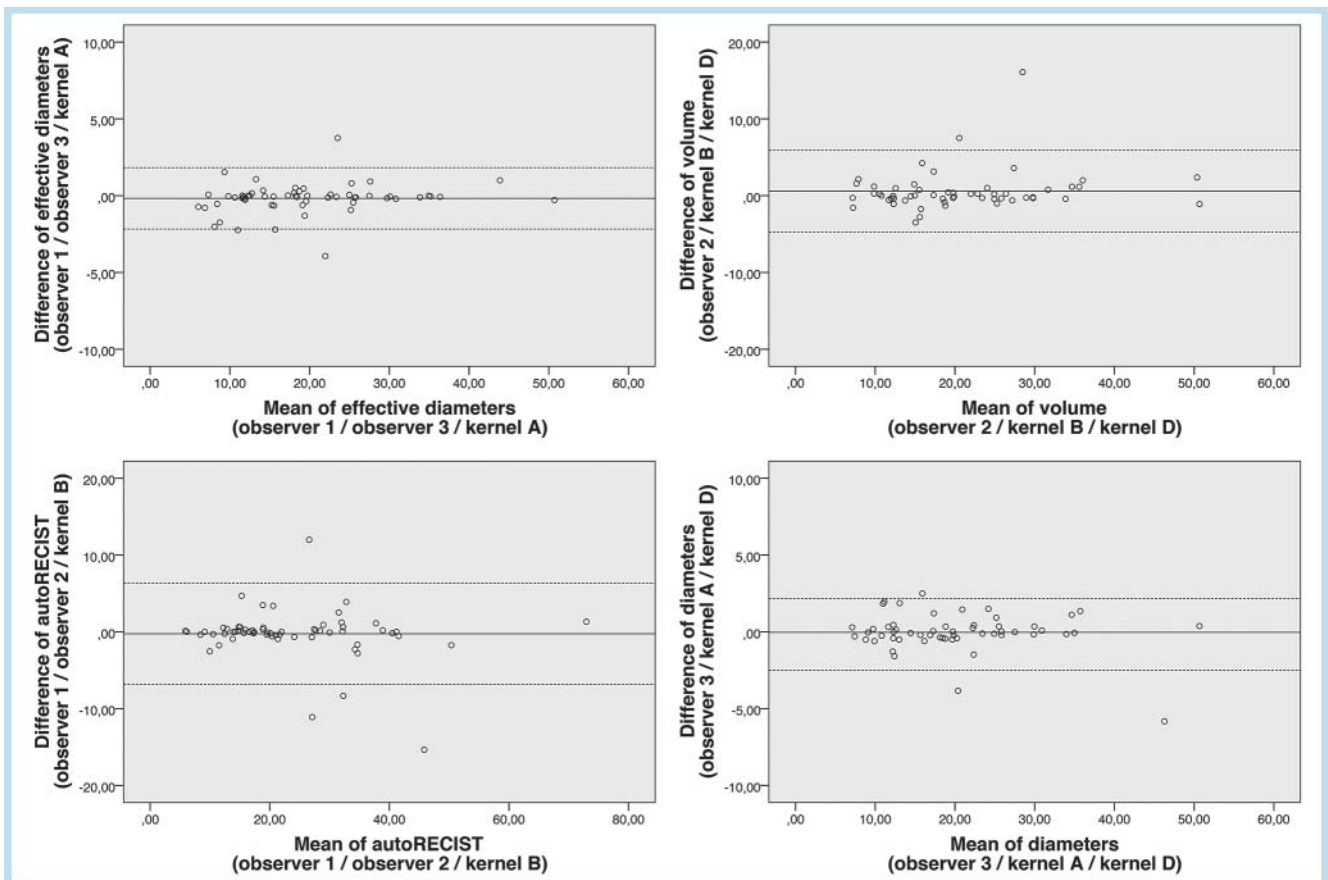


Fig. 2 Examples of Bland-Altman plots for different correlation analyses. The plots show the mean differences in the respective measurements and the corresponding limits of agreement (mean difference ± 1.96 standard deviation) for each comparison. Similar plots were obtained for all correlations.

Abb. 2 Beispiele von Bland-Altman-Plots für verschiedene Korrelationsanalysen. Die Plots zeigen die mittlere Differenz der jeweiligen Messungen und die entsprechenden Limits of Agreement (mittlere Differenz $\pm 1,96$ Standardabweichung). Ähnliche Plots ergaben sich für alle Korrelationen.

Discussion

The liver is one of the organs most affected by metastatic disease [19]. Until today, guidelines for staging and monitoring therapeutic response are based on unidimensional measurements which have been shown to be inconsistent to some extent. They do not account for anisotropic tumor growth which might lead to different assessment of tumor response [4, 20, 21]. Recent studies support the evidence that volumetric measurements might improve the assessment of tumor growth [22, 23]. Still, until today there are no guidelines for assessing tumor growth and therapy response through segmentation and volumetry.

When using semi-automated segmentation, attention must be paid to the different parameters of MDCT datasets used for measurements as they might influence the results. In a phantom study with relatively big lesions, the slice thickness had no influence on the measurement results [15]. However for the measurement of liver lesions under more realistic conditions in vivo, Puesken et al. showed that the use of a greater slice thickness impacts the precision of semi-automated measurements [14]. In the clinical routine and retrospective oncological trials not only slice thickness may vary but also different scanners and different reconstruction kernels might be used. Again, in a phantom study different kernels showed no significant influence [15]. In our study we aimed at evaluating the effect of different reconstruction kernels for the measurement of liver lesions in vivo. To eliminate the influence

of manual correction, only lesions with satisfactory initial segmentation results were included. To further reduce the user's influence on segmentation results, measurements in other than the standard kernel were performed by exactly copying segmentation start parameters using the software's dedicated tool thus preventing intraobserver variability. Of course, this is also a limitation of the study to some extent, as the initial seed diameter could have been drawn differently in other kernels possibly resulting in a different segmentation result. Although we only analyzed a limited number of lesions, our data shows that for the in vivo situation no significant influence of different reconstruction kernels on semi-automated segmentation can be found, nor is there a trend. ICCs showed a very high concordance (≥ 0.96) for both interobserver and interkernel variability. Differences in the segmentation results were independent of lesion size (• Fig. 2). A partial volume effect has a greater influence in smaller lesions and could vary considerably in different kernels. We restricted the evaluation of lesions > 5 mm since the status and the clinical relevance of such lesions is questionable. Accordingly, only lesions ≥ 10 mm are defined as measurable according to mRECIST. Today the gold standard is unidimensional manual measurement [24]. For the semi-automated unidimensional measurements, the mean difference between different reconstruction kernels ranged around 1% with a maximum standard deviation of around 7%. Of course differences are more evident for the volumetric measurements as it is the third power of a unidimensional

radius. Thus, we had a maximum mean difference of around $5 \pm 25\%$ which however did not reach statistically significant levels. Therefore, we conclude that semi-automated 3D-volumetric analysis allows for reliable measurements of liver lesions regardless of the kernel used for reconstruction of the MDCT dataset.

References

- 1 Eisenhauer EA *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; 45: 228–247
- 2 Therasse P. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *et al.* New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst* 2000; 92: 205–216
- 3 Marten K *et al.* Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria. *Eur Radiol* 2006; 16: 781–790
- 4 Prasad SR *et al.* CT tumor measurement for therapeutic response assessment: comparison of unidimensional, bidimensional, and volumetric techniques initial observations. *Radiology* 2002; 225: 416–419
- 5 Bornemann L *et al.* OncoTREAT: a software assistant for cancer therapy monitoring. *International Journal of Computer Assisted Radiology and Surgery* 2007; 1: 231–242
- 6 Fabel M *et al.* Semi-automated volumetric analysis of lymph node metastases in patients with malignant melanoma stage III/IV—a feasibility study. *Eur Radiol* 2008; 18: 1114–1122
- 7 Dornheim J *et al.* Segmentation of neck lymph nodes in CT datasets with stable 3D mass-spring models segmentation of neck lymph nodes. *Acad Radiol* 2007; 14: 1389–1399
- 8 Wessling J *et al.* MSCT follow-up in malignant lymphoma: comparison of manual linear measurements with semi-automated lymph node analysis for therapy response classification. *Fortschr Röntgenstr* 2012; 184: 795–804
- 9 Das M *et al.* Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners. *Eur Radiol* 2007; 17: 1979–1984
- 10 Heussel CP *et al.* Quantitative CT-Verlaufskontrolle von Lebermalignomen nach RECIST und WHO im Vergleich zur Volumetrie. *Fortschr Röntgenstr* 2007; 179: 958–964
- 11 Piesken M *et al.* Einfluss des vaskularisationsgrades auf die automatische segmentierung und messung von lebertumoren nach RECIST in einer biphasischen multi-slice-CT (MSCT). *Fortschr Röntgenstr* 2009; 181: 67–73
- 12 Wulff AM *et al.* Lung, liver and lymph node metastases in follow-up MSCT: comprehensive volumetric assessment of lesion size changes. *Fortschr Röntgenstr* 2012; 184: 820–828
- 13 Kalkmann J *et al.* Suitability of semi-automated tumor response assessment of liver metastases using a dedicated software package. *Fortschr Röntgenstr* 2010; 182: 581–588
- 14 Piesken M *et al.* Liver lesion segmentation in MSCT: effect of slice thickness on segmentation quality, measurement precision and interobserver variability. *Fortschr Röntgenstr* 2011; 183: 372–380
- 15 Keil S *et al.* Semi-automated quantification of hepatic lesions in a phantom. *Invest Radiol* 2009; 44: 82–88
- 16 Moltz JH *et al.* Advanced Segmentation Techniques for Lung Nodules, Liver Metastases, and Enlarged Lymph Nodes in CT Scans. *Selected Topics in Signal Processing, IEEE J-STSP* 2009; 3: 122–134
- 17 Kwiekien R, Kopp-Schneider A, Blettner M. Concordance analysis: part 16 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2011; 108: 515–521
- 18 Koch R, Spörl E. Statistische Verfahren zum Vergleich zweier Messmethoden und zur Kalibrierung: Konkordanz-, Korrelations- und Regressionsanalyse am Beispiel der Augeninnendruckmessung. *Klin Monbl Augenheilkd* 2007; 224: 52–57
- 19 Koshariya M *et al.* An update and our experience with metastatic liver disease. *Hepatogastroenterology* 2007; 54: 2232–2239
- 20 Harris KM *et al.* The effect on apparent size of simulated pulmonary nodules of using three standard CT window settings. *Clin Radiol* 1993; 47: 241–244
- 21 Erasmus JJ *et al.* Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J Clin Oncol* 2003; 21: 2574–2582
- 22 Zhao B *et al.* A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development. *Clin Cancer Res* 2010; 16: 4647–4653
- 23 Liu F *et al.* Assessment of therapy responses and prediction of survival in malignant pleural mesothelioma through computer-aided volumetric measurement on computed tomography scans. *J Thorac Oncol* 2010; 5: 879–884
- 24 Jaffe CC. Measures of response: RECIST, WHO, and new alternatives. *J Clin Oncol* 2006; 24: 3245–3251