Comparison of Volumetric and Linear Serial CT Assessments of Lung Metastases in Renal Cell Carcinoma Patients in a Clinical Phase IIB Study

Volker Dicken, PhD, Lars Bornemann, MS, Jan Hendrik Moltz, PhD, Heinz-Otto Peitgen, PhD, Souhil Zaim, MD, Urban Scheuring, MD

Rationale and Objectives: Accuracy of radiologic assessment may have a crucial impact on clinical studies and therapeutic decisions. We compared the variability of a central radiologic assessment (RECIST) and computer-aided volume-based assessment of lung lesions in patients with metastatic renal cell carcinoma (RCC).

Materials and Methods: The investigation was prospectively planned as a substudy of a clinical randomized phase IIB therapeutic trial in patients with RCC. Starting with the manual study diameter (SDM) of the central readers using RECIST in the clinical study, we performed computer-aided volume measurements. We compared SDM to an automated RECIST diameter (aRDM) and the diameter of a volume-equivalent sphere (effective diameter [EDM]), both for the individual size measurements and for the change rate (CR) between consecutive time points. One hundred thirty diameter pairs of 30 lung lesions from 14 patients were evaluable, forming 55 change pairs over two consecutive time points each.

Results: The SDMs of two different readers showed a correlation of 95.6%, whereas the EDMs exhibited an excellent correlation of 99.4%. Evaluation of CRs showed an SDM-CR correlation of 63.9%, which is substantially weaker than the EDM-CR correlation of 87.6%. The variability of SDM-CR is characterized by a median absolute difference of 11.4% points versus the significantly lower 1.8% points EDM-CRs variability (aRDM: 3.2% points). The limits of agreement between readers suggest that an EDM change of 10% or 1 mm can already be significant.

Conclusions: Computer-aided volume-based assessments result in markedly reduced variability of parameters describing size and change, which may offer an advantage of earlier response evaluations and treatment decisions for patients.

Key Words: RECIST; MDCT; lung nodule; volumetry.

Clinical oncologic development studies have sought to improve the precision of radiology assessments and to reduce the inter-reader variability impacting on study endpoints such as time to progression (TTP), progression-free survival (PFS), or response rate by using a central blinded radiology review. Although this standardized approach entailing two-skilled readers and an adjudicator in case of divergent results has increased the data quality, further improvement may be of high interest. It has been documented that intraobserver and interobserver variability of radiologic assessment (RECIST) (1) measurements of lung lesions is considerable and may lead to significant misclassifications (2). However, it has also been shown that computer-aided volumetric assessment of lung nodules may allow a reliable classification as progressive disease (PD) at a volume increase of 27%, as opposed to 73% (73% volumetric increase correspond to 20% increase in the sum of longest diameters) that is required by RECIST (3).

In this study, we investigate in more detail how computer-aided assessment of lung lesions reduces the measurement variability compared to manual measurements of in-plane

Acad Radiol 2015; 22:619–625

From the Department of Image Processing, Fraunhofer MEVIS, Universitaetsallee 29, D-28359 Bremen, Germany (V.D., L.B., J.H.M., H.O.P.); Department of Radiology, BioClinica, Newark, CA, USA (S.Z.); Department of Oncology/Infectiology, Bayer Vital GmbH, BV-PH-MED-ONC/INF, Leverkusen, Germany (U.S.). Received August 5, 2014; accepted December 23, 2014. Funding Sources: This study has been supported financially by BayerHealthCare. **Address correspondence to:** V.D. e-mail: volker.dicken@mevis.fraunhofer.de

http://dx.doi.org/10.1016/j.acra.2014.12.018

lesion diameters as performed in clinical trials according to RECIST. Furthermore, we put a special focus on how this reduction translates to the variability of change rates (CRs) computed from follow-up measurements and whether the RECIST thresholds could be adapted.

MATERIAL AND METHODS

Material

Our study is based on data from an oncologic clinical trial that studied sorafenib versus IFN-2a as a first-line treatment in patients with advanced or metastatic renal cell carcinoma (RCC) (4). The primary collection of study data consists of the full set of computed tomography (CT) images and annotations from the central RECIST reading of the clinical trial including 174 patients and 726 CT studies. CT scans were taken after every other cycle of therapy; some patients got intermediate scans because of medical need or when they discontinued treatment. The data were assessed by two readers independently. A third reader (adjudicator) was only asked to read data from a subject if readers 1 and 2 did not agree at a patient level with regard to the RECIST classification. The RECIST system provides a classification for changes in the sum of longest diameters of target tumor lesions. It comprises the categories complete response with disappearance of all tumor lesions, partial response (PR) for tumor reductions by >30%, stable disease (SD) for tumor decreases by ≤30% or tumor increases by ≤20%, and PD for tumor increases by >20% according to certain criteria (1). Although these RECIST thresholds are not applied on a per lesion basis, but on the sum of diameters of all target lesions, we used the RECIST thresholds on a per lesion basis in the present study as we usually evaluated only a few selected lung lesions per patient that were 3D-measurable and measured by two or more readers at multiple time points.

All manual diameter measurements were archived. The central radiologic reading recorded 11,659 findings in total.

For our analysis, we extracted all findings labeled with LUNG for which high-resolution images were available from at least three time points. High resolution was defined by a slice increment of at most 4 mm to allow for volumetric measurements. For each of these findings, the diameters as originally recorded in the clinical trial were used to provide the manual study diameter (SDM) measurements and to initialize semiautomatic segmentation using an algorithm described by Kuhnigk et al. (5). The algorithm requires an approximate lesion diameter as input to generate a cuboid region of interest (ROI) on which an automated segmentation algorithm segments the lesion. It mainly involves region growing and adaptive morphologic image processing operations (smart opening) and a partial volume analysis to compute the lesion volume. The diameters originally drawn by the study reader on the lung finding were used to initialize the lesion segmentation, provided the high-resolution CT data required for volumetry were in the same breathing position as the scan on which the finding was marked (often study findings were marked on traditional 5-mm increment CT data not on the high resolution data.) If the breathing position differed strongly, the software operator had to draw a similar approximate diameter at the shifted lesion position to start the segmentation. Obvious segmentation errors were corrected interactively by the operators using the tools by Heckel et al. (6). Screenshots featuring 3 × 9 cross-sections through lesion ROI (9 for each axial sagittal and coronal orientation) were generated for each segmented lesion and checked visually. If the segmentation was not acceptable (eg, very inconsistent at different time points), the related results were removed from the statistical evaluation. We also excluded some cases where the difference between manually defined study diameters exceeded 40%, considering these lesions as nonmeasurable because multiple readers were not able to agree on a boundary.

From the software results, two measurements were taken into account for further analysis, the automated RECIST diameter (aRDM) defined as the maximum in-plane distance of two voxels inside a segmentation mask provided by the software, and the effective diameter (EDM) defined as the diameter of a sphere having the volume determined by the software for the lesion.

$$\mathrm{EDM} := \left(\frac{6}{\pi}\mathrm{Volume}\right)^{1/3}$$

To analyze the variability of measurements and derived CRs, we restricted our study to lesions that had been measured by at least two readers at multiple time points and where a volumetric measurement was available at all time points. This left 285 individual findings, belonging to 30 lesions in 14 patients, measured at three to nine time points by two or three readers.

Methods

We compared three measurements of lesion size:

1) SDM: Manually defined maximum in-plane diameter according to RECIST criteria as drawn during central reading by the imaging core lab radiologists.
2) aRDM: Automatically computed maximum in-plane diameter.
3) EDM: Diameter of a sphere with the same volume as computed automatically for the lesion.

The reason for using the EDM is that the volume itself cannot be compared directly to a diameter. Partial volume effects were taken into account as described in (5) when calculating a lesion's volume, as it was shown to significantly reduce variations.

CRs are computed between each pair of consecutive time points independently, according to the formula, $CR(t_{n+1}, t_n) = \frac{DM(t_{n+1}) - DM(t_n)}{DM(t_n)}$. For most cases, high-resolution baseline data were not available, and we aimed to compute CRs over comparable time intervals.