Radiotherapy and Oncology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original article

Prostate tumor delineation using multiparametric magnetic resonance imaging: Inter-observer variability and pathology validation

Peter Steenbergen [a], Karin Haustermans [c], Evelyne Lerut [e], Raymond Oyen [d], Liesbeth De Wever [d], Laura Van den Bergh [c], Linda G.W. Kerkmeijer [f], Frank A. Pameijer [g], Wouter B. Veldhuis [g], Jochem R.N. van der Voort van Zyp [f], Floris J. Pos [a], Stijn W. Heijmink [b], Robin Kalisvaart [a], Hendrik J. Teertstra [b], Cuong V. Dinh [a], Ghazaleh Ghobadi [a], Uulke A. van der Heide [a,*]

[a] Department of Radiation Oncology; [b] Department of Radiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands; [c] Radiation Oncology, University Hospitals Leuven, Department of Oncology, KU Leuven; [d] Radiology; [e] Pathology, University Hospitals Leuven, Department of Imaging & Pathology, KU Leuven, Belgium; [f] Department of Radiation Oncology; and [g] Department of Radiology, University Medical Center Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 13 October 2014
Received in revised form 27 March 2015
Accepted 19 April 2015
Available online xxxx

Keywords:
Prostate cancer
Radiotherapy
Tumor delineation
Multiparametric MRI
Dose escalation
Inter-observer variability

ABSTRACT

Background and purpose: Boosting the dose to the largest (dominant) lesion in radiotherapy of prostate cancer may improve treatment outcome. The success of this approach relies on the detection and delineation of tumors. The agreement among teams of radiation oncologists and radiologists delineating lesions on multiparametric magnetic resonance imaging (mp-MRI) was assessed by measuring the distances between observer contours. The accuracy of detection and delineation was determined using whole-mount histopathology specimens as reference.

Material and methods: Six observer teams delineated tumors on mp-MRI of 20 prostate cancer patients who underwent a prostatectomy. To assess the inter-observer agreement, the inter-observer standard deviation (SD) of the contours was calculated for tumor sites which were identified by all teams.

Results: Eighteen of 89 lesions were identified by all teams, all were dominant lesions. The median histological volume of these was 2.4 cm^3. The median inter-observer SD of the delineations was 0.23 cm. Sixty-six of 69 satellites were missed by all teams.

Conclusion: Since all teams identify most dominant lesions, dose escalation to the dominant lesion is feasible. Sufficient dose to the whole prostate may need to be maintained to prevent under treatment of smaller lesions and undetected parts of larger lesions.

Boosting the radiation dose to the visible cancer inside the prostate may improve treatment outcome. This theory is motivated by the observation that prostate cancer often recurs at the site of the dominant lesion [1,2] and by observations that escalating the radiation dose improves treatment outcome [3]. A boost to the largest lesion in the prostate (the dominant lesion), in addition to a standard dose to the whole organ may achieve adequate dose escalation without increasing toxicity [4,5]. This hypothesis is currently tested in a phase III clinical trial, the FLAME trial [6]. Boost volumes can be defined on multiparametric magnetic resonance imaging (mp-MRI).

Mp-MRI, consisting of a combination of T2-weighted (T2w), diffusion-weighted (DWI) and dynamic contrast enhanced (DCE) images, has a high diagnostic accuracy for detection of prostate cancer [7,8]. By dividing the gland into regions and scoring tumor presence per region, specificities and sensitivities for tumor detection in the range of 0.53–0.81 and 0.80–0.96 are feasible [9]. Sensitivity using all three modalities together is higher than for the separate modalities [10].

The inter-observer agreement of tumor detection can be quantified using a kappa statistic. The number of volume elements (regions of the prostate or voxels in the images) for which the observers agree is expressed as a value ranging from zero (no agreement) to one (perfect agreement). Studies of region-based tumor detection on mp-MRI have reported kappa values ranging from 0.40 to 0.63 [11,12]. Rischke et al. [13] performed a kappa analysis on a voxel-by-voxel basis in five patients. They compared tumor delineations of five observers with delineations by a reference observer, resulting in an agreement of 0.51 ± 0.15 (mean ± standard deviation; range 0.22–0.73) for a combination of T2w and DWI images and 0.63 ± 0.12 (range 0.00–0.80) for T2w with DCE images. Delineation on a combination of all three

* Corresponding author at: The Netherlands Cancer Institute, Department of Radiation Oncology, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.
E-mail address: u.vd.heide@nki.nl (U.A. van der Heide).

http://dx.doi.org/10.1016/j.radonc.2015.04.012
0167-8140/© 2015 Published by Elsevier Ireland Ltd.

MR sequences was not studied by these authors. Anwar et al. compared delineations by two observers on T2w imaging and MR spectroscopy with delineations on histology and found these to differ by a median distance of 1.4 mm [14].

In this study, we determine the inter observer agreement and geometrical distances between the contours of the observers. The latter is the measure closest to the delineation practice in radiotherapy. Six teams from three different centers were recruited, each team consisting of a radiation oncologist and a radiologist, which delineated prostate tumors on mp-MRI. By comparing the results with histology, we assessed to what extent tumors were correctly identified and what the agreement was between the observer and pathology contours.

Methods

Dataset

Twenty patients with biopsy proven prostate cancer were selected who consulted the Department of Radiation Oncology of the University Hospitals Leuven. The patients were prospectively included in an earlier study [10,15]. Between February 2008 and February 2011 they underwent an mp-MRI exam before prostatectomy. The patients were selected such that the majority had T2c or T3a prostate cancer based on analysis of the prostatectomy specimens. Apart from tumor stage, the selection was random. Written informed consent was obtained from all patients.

The mp-MRI scan consisted of T2-weighted (T2w), diffusion weighted (DWI), and dynamic contrast-enhanced (DCE) scans. These were acquired on 1.5T MR scanner (SonataVision, SymphonyVision or Aera, Siemens Erlangen, Germany) using a combination of a six-channel phased-array body coil and a spine coil. Because the patients were selected retrospectively, the scanning parameters varied (see Supplementary material). Orthogonal transversal, coronal and sagittal T2w scans were acquired. ADC maps were calculated from the DWI scan by the scanner software. For the DCE scans, the contrast agent (Dotarem, Guerbet, France, 15 ml) was injected at the fourth dynamic scan with a rate of 2 ml/s, followed by a 20 ml saline flush. The signal intensities were first converted to gadolinium concentration values [16] using reference T1 values [17]. The extended Tofts model was fitted to the concentration time curves for estimation of the volume transfer constant (K^{trans}) using the method by Murase [18] with a population-based arterial input function.

Six teams of delineators were recruited from three hospitals (two teams per hospital). Each team consisted of a radiation oncologist and a radiologist. Using the mp-MRI, they delineated the visible tumors on the transversal T2 (T2t) scan. To assure that the study results are representative for a realistic clinical situation, the teams used the delineation system of their own hospital. The teams were provided with biopsy information, TRUS findings and written radiology reports. The final delineations were approved by both members of the team.

Pathology

Whole-mount histological slices were cut from the prostatectomy specimens perpendicular to the urethra at 3–4 mm intervals [10,15]. From these, Hematoxylin-eosin (H&E) stained whole-mount pathology slides were obtained on which a single pathologist delineated the tumors. The H&E slides were registered to the T2t image to allow assessment of tumor detection accuracy of the observers. First, each pathology slide was assigned to a T2t slice based on the relative order of the H&E slides, the location of apex and base, size of the subsequent slices in MRI and pathology and the distance between the slides (3 or 3.3 mm in T2t, 3–4 mm in histology). For some T2t slices no H&E slice was present due to the difference in slice distance. The location of the pathology slides does not necessarily coincide with the location of the MRI slices. Therefore the true location of an H&E slice in the T2t scan can be somewhere between two slices. The T2t slice distance is 3.3 mm for most patients; assuming that the best matching T2w slice is selected, the average error is 0 mm, ranging from −1.65 to +1.65 mm. The average absolute error in the slice direction is then 0.8 mm.

Subsequently, each slice was registered by means of a deformable point-based method (Coherent Point Drift) using landmark points which were visible on both images. These were mostly based on the prostate boundary and occasionally other features, such as the ends of the peripheral zone and transitions between prostate and seminal vesicles. We estimated the registration error by selecting one landmark not used for registration per pathology slice and measuring the distance between this point on T2t MRI and registered pathology. The average error we found was 2.1 mm; the largest error was 5 mm. We could not assess the error resulting from possible differences in orientation of the pathology slices and the imaging.

The pathologist delineated the tumor locations on the H&E slices prior to registration. These delineations were digitized after image registration. The contours of seven dominant lesions and one satellite were interpolated for slices where no matching H&E slice was present while a tumor was found at the same site in both the previous and next slice.

Analysis

All contour analyses were done by one author using each patient's T2t scan grid as unit of analysis. The contours of the same site by different teams were grouped manually based on overlap. Depending on whether or not a corresponding pathology contour was found in the histology, the contour groups were counted as true positives, false positives and false negatives. Delineations made on MRI slices without matching histology were categorized as unconfirmed.

The prostate voxels were divided in healthy/tumor based on pathology and in true/false positive/negative based on pathology and consensus detections by all teams. The median and quartile of the voxel values of these categories were calculated for all scans. For this, all images were first normalized to their median.

The volumes of the observer delineations were estimated by counting the number of T2t voxels in the contours and multiplying this by the voxel volume. The same was done for the digitized pathologist's contours. Differences between teams in delineated volumes were analyzed using the Friedman test in Matlab (version 8.1, Statistics Toolbox 8.2, Natick, Massachusetts: The MathWorks Inc., 2013), with the observer teams as factor and patients as "nuisance factor".

The inter-observer delineation variation of the tumor sites which were delineated by all teams was assessed relative to a reference contour. The measure of interest was the standard deviation of the contour distances. The exact shape of the reference contour is not important for this, as long as it is smooth and similar in shape to the contours to be tested. The pathologist's contours had irregular shapes in some cases, making them unsuitable as a reference. Instead, the median of the observer contours was used, which was obtained by converting each contour to a mask, adding the six masks, smoothing the result and contouring the 50% level. The perpendicular distances of all points of the reference contour to each observer contour were then calculated and pooled [19]. The standard deviation (SD) of this pool was then calculated to obtain the inter-observer variability for one tumor site.

To facilitate comparison of the study results with literature, we assessed the agreement at the voxel level using kappa indices.

Please cite this article in press as: Steenbergen P et al. Prostate tumor delineation using multiparametric magnetic resonance imaging: Inter-observer variability and pathology validation. Radiother Oncol (2015), http://dx.doi.org/10.1016/j.radonc.2015.04.012