Validation and Comparison of Pre- and Intraoperative Repeated Image Segmentation in MRI-Guided Brachytherapy Based on Spatial Overlap Statistics

K. H. Zou[1,2], A. Bharatha[3], N. Hata[4], M. Hirose[5], S. J. Haker[1], R. A. Cormack[6], W. M. Wells[1,7], A. D'Amico[6], R. Kikinis[1], F. A. Jolesz[1], C. M. Tempany[1]

[1]Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States, [2]Health Care Policy, Harvard Medical School, Boston, MA, United States, [3]Medical Imaging, Univerisity of Toronto Medical School, Toronto, Ontario, Canada, [4]Mechano-Informatics, University of Tokyo, Tokyo, Japan, [5]Radiology, Showa University, Tokyo, Japan, [6]Radiation Oncology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States, [7]Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Boston, MA, United States

INTRODUCTION.

The performance of segmentation methods has a direct impact on detection and target definition in MR-guided interventions. The clinical goal of surgical planning and quantitative monitoring of disease progression requires highly reproducible segmentation due to limited number of images per case. We aim to evaluate the performance of repeated manual segmentation of preoperative 1.5T and intraoperative 0.5T MR images of the prostate's peripheral zone (PZ) collected before and during brachytheraphy [1-3]. Without a known gold standard obtained by non-imaging methods such as histology, the validation task becomes an assessment of reliability or reproducibility. A simple spatial overlap index is the Dice similarity coefficient (DSC), a spatial overlap index and a reproducibility validation metric. The value of a DSC ranges from 0 indicating no spatial overlap between two sets of binary segmentation results, to 1 indicating complete overlap [4].

MATERIALS AND METHODS.

Subjects: A total of 10 sequential MR-guided brachytherapy cases were identified retrospectively, excluding those who had received any prior brachytreatment or previous external beam radiation therapy, which could confound shape and MR signal intensity changes in the gland [1]. Imaging: (1) Preoperative: All patients underwent preoperative 1.5T-MR imaging using an endorectal coil (MedRad, Indianola, PA) with an integrated pelvic-phased multicoil array (Signa LX, GE Medical Systems, Milwaukee, WI). The endorectal coil is a receive-only coil mounted inside a latex balloon, and assumed a diameter of 4-6 cm, once inflated in the patient's rectum. The patient was placed in a supine position in the closed-bore magnet for the imaging examination. The axial T2-weighted images were fast spin echo (FSE) images (4050/135, field of view of 12 cm, section thickness of 3 mm, section gap of 0 mm, matrix of 256x256, 3 signal averages). Typical acquisition times were 5-6 min. (2) Intraoperative: Imaging was performed in the open-configuration 0.5T MR scanner (Signa SP, General Electric Medical Systems, Milwaukee, WI). Each patient was placed in the lithotomy position in order to facilitate prostate brachytherapy via a perineal template. The perineal template was fixed in place by a rectal obturator (2 cm in diameter). T2-weighted FSE images (axial and coronal, 6400/ 100, field of view of 24 cm, section thickness of 5 mm, section gap of 0 mm, matrix of 256x256, 2 signal averages) were acquired in the MR scanner using a flexible external pelvic wraparound coil, with typical acquisition times of 6 min. Image Segmentations: The 3D Slicer (http://www.slicer.org) was used as a surgical simulation and navigation tool in order to facilitate manual segmentation. Manual contouring of two areas of the prostate, the PZ and the central gland, was performed by two segmenters, using the T2-weighted images from the 1.5 and 0.5 T studies. The PZ of the prostate is the clinical target volume for brachytherapy in the clinical practice. Thus, each segmenter independently and blindly outlined the PZ, in randomly selected 5 of these 10 cases. The manual segmentations of the same preoperative 1.5T and intraoperative 0.5T image were also repeated 5 times. When segmenting the intraoperative 0.5T images, the segmenters were allowed to examine preoperative 1.5T images, as is done in clinical practice. Statistical Methods: We compared the the reproducibility of segmenting the PZ based on pre- vs. intraoperative images using all pairwise spatial overlaps of 5 repeated segmentations of the PZs of the 10 brachytherapy cases pooled over the 2 image segmenters as we did not find significant differences between these segmenters. We tested whether there existed a learning-curve effect over time. We labeled a segmentation pair as S_k and $S_{k'}$ from all 10 parings of the 5 repeated segmentations. For each segmentation pair, summary statistics including the means and the standard deviations (SD) of DSC and logit(DSC) were computed over all 10 cases, separately for preoperative 1.5T and intraoperative 0.5T images. Analysis of Variance (ANOVA) was conducted with the components labeled as M=preoperative 1.5T or intraoperative 0.5T MR, C=case, S=segmentation pair, and two possible interactions, S×M and S×C. Correspondingly, there were 2 imaging×10 cases×10 segmentation pairs=200 DSC values based on all pairwise overlaps of the repeated segmentations. In our ANOVA model, the outcome variable was logit(DSC) with a normal error assumption. Regression equation for logit-transformed DSC was: logit(DSC)= ln{DSC/(1-DSC}=μ+M+C+S+ S×M+ S×C+e, where normality was assumed for each component in the model; μ was the intercept and e is the residual error term. The F statistic and p-value were computed based on mean DSC as a validation metric [4]. All statistical analyses were performed using the software package, S-Plus (http://www.insightful.com).

RESULTS.

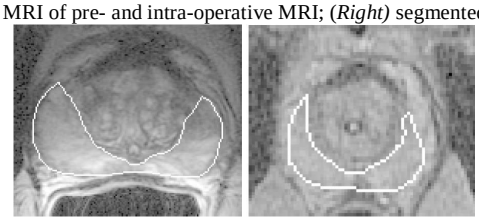
Summary Statistics: The preoperative and intraoperative images and intraoperative 0.5T T2-weighted MR images of a case are displayed, respectively (Figures). One of the five repeated manual contours of the PZ on these images is presented in each figure. The means DSCs values of all segmentation pairs based on both preoperative 1.5T and intraoperative 0.5 images are given (Table). The pre- and intraoperative imaging resulted in significantly higher reproducibility by mean logit (DSC) preoperatively than intraoperatively The mean DSCs of the 10 cases were 0.883 (range 0.876-0.893) preoperatively and 0.838 (range 0.819-0.852) intraoperatively ($p<0.001$) (Tables). Pairwise logit-transformed of the 10 repeated segmentations of each of the 10 cases yielded non-significant normality test results (range 0.27 to 0.81 on 1.5T and 0.07-0.80 on 0.5T). Comparing the mean logit(DSC) values, they were 2.070 (range 2.011-2.159) on 1.5T versus 1.659 (range 1.525-1.742) on 0.5T; thus, the segmentation reproducibility appeared higher based on preoperative images. ANOVA: According to all pairwise DSCs over segmentation pairs and the DSCs of only sequentially consecutive segmentation pairs (Tables), there was a statistically significant improvement in the reproducibility results using 1.5T over 0.5T images ($p<0.001$). We did not observe a significant effect of segmentations among all pairs ($p=0.12$), nor was there a learning curve phenomenon ($p=0.97$). A case-to-case variation was however statistically significant ($p<0.001$).

Tables: (*Left*) Mean DSC and (*Right*) ANOVA of effects of variance components.

Segmentation Pair	Pre Op 1.5T	Intra Op 0.5T
(1,2)	0.879	0.834
(2,3)	0.887	0.841
(3,4)	0.880	0.840
(4,5)	0.893	0.852

Variance Component	n	F	p-value
MR (Pre vs. Intra Op)	2	202.004	<0.001
Case	10	25.299	<0.001
Segmentation Pair	10	1.631	0.12
Segmentation×MR	10	0.381	0.94
Segmentation×Case	100	0.329	>0.99

Figures: (*Left*) MRI of pre- and intra-operative MRI; (*Right)* segmented PZs.



DISCUSSION.

The reproducibility was significantly higher based on preoperative 1.5T images than on intraoperative 0.5T images in brachytherapy. This necessitates developing a registration method to register to achieve improved visualization in brachtherapy [2]. There was no significant difference in mean logit(DSC) over segmentation pairs. Furthermore, repeated segmentations did not introduce significant bias.

REFERENCES.

[1] Hirose M et al. Acad Radiol. 2002; 9: 906-912.
[2] Bharatha A et al. Med Phys 2001; 28: 2551-2560.
[3] McTavish et al. ISMRM 1999, 1972.
[4] Zou et al. MICCAI 2002, LNCS 2488, 315-322.