Validation study of ESUR prostate MR guidelines 2012: the significance of different sequences of multi-parameter MRI in detection of prostate cancer in 106 consecutive patients

Juan Hu[1], He Wang[1], Xuedong Yang[1], Chengyan Wang[2], Rui Wang[1], Ge Gao[1], Hui Zhang[2], Wenchao Cai[1], Wei Wang[1], Jue Zhang[2,3], and Xiaoying Wang[1,2]

[1]Radiology Department, Peking University First Hospital, Beijing, Beijing, China, [2]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, Beijing, China, [3]College of Engineering, Peking University, Beijing, Beijing, China

Target audience:
Uroradiologists.

Purpose:
To evaluate the significance of each sequence of multi-parametric MRI (Mp-MRI) in detection of prostate cancer(PCa) based on PI-RADS scoring system of ESUR prostate MR guidelines 2012, and to find a appropriate combined mode of Mp-MRI to further improve the diagnostic efficacy of prostate MR imaging.

Materials and Methods :
Data source: A retrospective study was performed based on the database of 106 consecutive patients (40-85 years old, mean 67 yrs, PSA 0.30-297.30 ng/ml, mean 27.33 ng/ml), who were suspected of PCa by urologist. Mp-MRI was performed in all the patients with subsequent ultrasound guided biopsy within 3 months. All of them were underwent a long-term follow-up(18-31 months, mean 25 month). Fifty-three patients were confirmed to be PCa by biopsy or surgery, while the other 53 were not detected of PCa by serial biopsies and long-term follow-up. All Mp-MRI sequences included axial T2 weighted images (T2WI), diffusion weighted images (DWI) and MR spectroscopy (MRS). 55 of the data contained dynamic contrast enhanced images (DCE).

Image analysis: Two experienced radiologists (7 and 9 years of prostate MR experience) received two circles training of PI-RADS scoring following ESUR prostate MR guidelines 2012. Both radiologists, unknown of the histological diagnosis, graded the PI-RADS score of images of each sequence. Then they gave an overall score (5-point scale, 1-definitely benign; 2-probably benign; 3-indeterminate; 4-probably malignant; and 5-definitely malignant) of each patient when the clinical information (age, T-PSA, F/T PSA) was provided. The average score of both radiologists were calculated and compared with the clinical diagnosis at the end of follow-up. Inter-reader variability was assessed. The best weighting value of each sequence (T2WI, DWI, DCE and MRS) was calculated by Fisher's linear discriminant analysis (FLDA). The alternating free-response receiver operating characteristic (AFROC) method was used to analyze different modes of combined application of T2WI score, DWI score, DCE score, MRS score, highest score, sum score, overall score and weighting score. Differences in the areas under ROC curve (AUCs), sensitivity, specificity and accuracy were calculated at a statistical significance of $P < 0.05$.
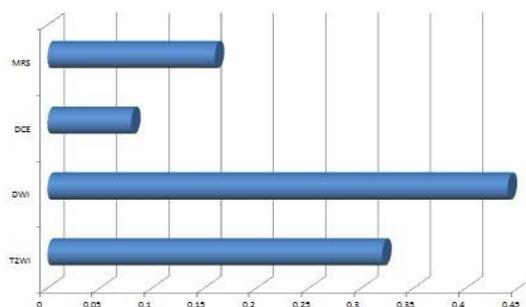

Fig.1. Fisher's linear discriminant analysis (FLDA) was used to calculate the best weighting factor of each MR sequence: DWI (0.42)>T2WI（0.35）>MRS（0.16）> DCE (0.08).
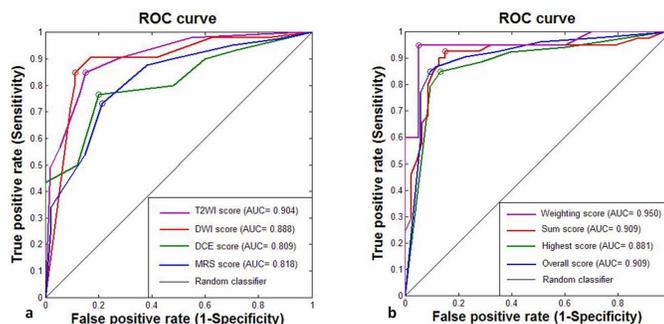
Results:
Inter-reader agreement was good (Kappa=0.692). The consistency of each sequence and overall score were different: overall score (0.830) > DWI (Kappa =0.754) > T2WI (Kappa=0.696) > DCE (Kappa=0.679) > MRS(Kappa=0.493). The weighting values obtained from FLDA were as follows: DWI (0.42) >T2WI (0.35)>MRS (0.16)> DCE (0.08)(Fig.1). The AUC, sensitivity, specificity and accuracy of weighting score were significantly higher than other modes of combined application($P < 0.05$). The best three of AUCs were weighting score (0.958)> overall score(0.909)≈ sum score(0.909). The three modes of best sensitivity were weighting score (0.950)> sum score(0.927)> overall score/T2WI score/DWI score/highest score(0.849). The three modes of best specificity were weighting score (0.950)> overall score (0.906)> DWI score (0.887). The three modes of best accuracy were weighting score (0.950)> sum score (0.886)> overall score (0.877)(Fig 2).

Discussion:
There was good sensitivity and specificity of PI-RADS score of each single MR sequence. While the Mp-MRI was combined by different mode, the diagnostic efficacy was elevated to varying degrees, especially the weighting score. The weighting values revealed that the most important sequences were DWI and T2WI, and more diagnostic information could be supplemented by MRS and DCE.

Conclusions:
PI-RADS scoring of ESUR prostate MR guidelines 2012 was an operable and useful scoring system for interpretation of Mp-MRI used in PCa detection. The diagnostic status of each MR sequence was not at the same level. Only if appropriate combined mode was used, the excellent diagnostic efficacy of Mp-MRI can be obtained. Furthermore, individual combined plan of Mp-MRI may be designed considering the balance of diagnostic efficacy, input of time and money.

Fig.2. ROC curves for **a)** single sequence: T2WI score(purple curve), DWI score(red curve), DCE score(green curve), MRS score(blue curve); **b)** different mode of combined application of

References:
[1]Jelle O, et al. ESUR prostate MR guidelines 2012 2012; 22:746-757.