

不同版本 PI-RADS(v1 与 v2)对前列腺癌诊断的一致性分析

居敏昊, 魏超刚, 陆志华, 赵文露, 沈钧康

【摘要】 目的:评价 PI-RADS v1 与 PI-RADS v2 在不同诊断经验医师间的评分结果一致性。方法:回顾性分析 175 例在 1.5T MR 设备行前列腺 T₂WI、DWI 和 DCEI 扫描并获得穿刺病理结果的前列腺癌患者影像资料。由两位具有不同诊断经验的医师分别应用 PI-RADS v1 和 v2 评分标准独立分析患者 mp-MRI 图像,记录两位医师 PI-RADS 评分结果。采用加权 Kappa 检验比较不同医师 PI-RADS 评分结果的一致性,并绘制 Bland-Altman 法画散点图作进一步对比分析。结果:在移行带病灶中,两位医师采用 PI-RADS v1 评分结果间的一致性为一般水平(二次加权 k 值为 0.337,95%可信区间 0.105~0.568),采用 PI-RADS v2 评分结果间的一致性处于中等水平(二次加权 0.498,95%可信区间 0.273~0.723);外周带病灶中,两位医师采用 PI-RADS v1 评分结果间的一致性为中等水平(二次加权 k 值 0.417,95%可信区间为 0.252~0.582),采用 PI-RADS v2 评分结果间的一致性处于良好水平(二次加权 k 值 0.658,95%可信区间 0.521~0.795);在不区分外周带及移行带病灶的综合评价中,两位医师 PI-RADS v1 评分结果一致性为中等水平(二次加权 k 值 0.398,95%的可信区间为 0.261~0.535),而两位医师 PI-RADS v2 评分结果一致性达到良好水平(二次加权 k 值 0.632,95%的可信区间为 0.510~0.755)。不同区域病灶的 Bland-Altman 散点图均表明在不同医师评分差值的平均值中,PI-RADS v2 较 PI-RADS v1 更接近于代表差值为零的线,提示无论在外周带还是移行带病灶的评价中,不同医师 PI-RADS v2 评分结果的一致性均高于 PI-RADS v1。结论:在 1.5T MR 设备和非直肠内线圈条件下,两位不同诊断经验医师采用 PI-RADS v2 对前列腺癌进行诊断评分的一致性高于 PI-RADS v1。

【关键词】 前列腺肿瘤;磁共振成像;诊断

【中图分类号】 R737.25; R445.2; R44 **【文献标识码】** A

【文章编号】 1000-0313(2019)10-1132-05

DOI:10.13609/j.cnki.1000-0313.2019.10.016

开放科学(资源服务)标识码(OSID):



The agreement research of two physicians for PI-RADS (v1/v2) diagnostic scoring at 1.5T MR equipment

JU Min-hao, WEI Chao-gan, LU Zhi-hua, et al. Department of Radiology, Xishan People's Hospital, Jiangsu 214000, China

【Abstract】 Objective: To preliminarily evaluate the inter-observer agreement between two physicians for diagnostic scoring of prostate cancer by applying PI-RADS v1 and PI-RADS v2. **Methods:** A total of 175 patients with biopsy confirmed prostate cancer were retrospectively included. T₂WI, DWI and DCE mp-MRI scans from 1.5T MRI scanner were reviewed and analyzed by two independent physicians according to the PI-RADS v1 and PI-RADS v2 scoring standard, and scored accordingly. The inter-observer agreement was tested by weighted Kappa method and the Bland-Altman plot. **Results:** In lesions of transition zone, the agreement of PI-RADS v1 score in different physicians was just fair (quadratic weighted k value was 0.337, 95%CI:0.105~0.568); the agreement of PI-RADS v2 score in different physicians was moderate (quadratic weighted k value was 0.498, 95%CI:0.273~0.723). In the peripheral zone, the agreement of PI-RADS v1 score in different physicians was moderate (quadratic weighted k value was 0.417, 95%CI:0.252~0.582), the agreement of PI-RADS v2 score in different physicians was good (quadratic weighted k value was 0.658, 95%CI:0.521~0.795). In all lesions,

作者单位:214000 江苏,无锡市锡山人民医院影像科/东南大学附属中大医院无锡分院(居敏昊);215004 江苏,苏州大学附属第二医院影像科(魏超刚、赵文露、沈钧康);215500 江苏,常熟市第一人民医院影像科(陆志华)

作者简介:居敏昊(1979-),男,江苏无锡人,硕士研究生,主治医师,主要从事腹部疾病的影像诊断和研究工作。

通讯作者:赵文露, E-mail: wenluzhao81@163.com

基金项目:苏州市科技发展计划(SS201534);苏州大学附属第二医院优势临床学科群(XKQ2015009)

the agreement of PI-RADS v1 score in different physicians was just moderate (quadratic weighted k value was 0.398, 95% CI: 0.261~0.535); the agreement of PI-RADS v2 score in different physicians was good (quadratic weighted k value was 0.632, 95% CI: 0.510~0.755). Draw the Bland-Altman scatter plot of the two physicians in diagnostic scoring of peripheral and transition zone cancer by using the PI-RADS v1 and PI-RADS v2 respectively, the results show that the mean of difference values in different physicians, PI-RADS v2 is closer to represent difference of zero line than PI-RADS v1, also revealed that considering the inter-reader agreement of PI-RADS score results in peripheral and transition zone cancer, PI-RADS v2 were higher than PI-RADS v1. **Conclusion:** At 1.5T MR scanner without the rectum coil. The inter-observer agreement of the two physicians by using PI-RADS v2 in prostate cancer was higher than PI-RADS v1.

【Key words】 Prostatic neoplasms; Magnetic resonance imaging; Diagnosis

欧洲泌尿放射学会(European Society of Urogenital Radiology, ESUR)在 2012 年推出了前列腺影像报告和数据库系统(prostate imaging reporting and data system, PI-RADS)^[1],并于 2014 年与美国放射学会(American College of Radiology, ACR)和 AdMeTech 基金会通过对 PI-RADS 的补充、完善和删减推出了第二版 PI-RADS^[2]。鉴于 PI-RADS v1 与 PI-RADS v2 的评分细节和流程存有异同,本文作者尝试在 1.5T MR 设备和非直肠内线圈条件下,探讨两位不同诊断经验医师分别应用不同版本 PI-RADS (v1、v2)对前列腺癌进行诊断时的评分结果的一致性。

材料与方法

1. 一般资料

回顾性分析 2009 年 12 月—2015 年 12 月 358 例行 1.5T 前列腺 mp-MRI 扫描(包括轴面高分辨率 T₂WI、DWI 和 DCEI)且经病理证实患者的影像资料,患者在接受 MR 检查前均未作相关针对前列腺疾病的治疗(如冷冻、手术、放疗和内分泌等)。所有患者均在 MR 检查后 6 周内行经直肠超声(transrectal ultrasound, TRUS)引导下前列腺系统穿刺活检。将前列腺由上到下平均分为基底部、体部、尖部 3 个部分,各部分平均分为左右 2 区,基底部、体部左右 2 区再均分为内外 2 部分,每区各行 1 针穿刺,并在可疑区域加穿 2~4 针。由泌尿系统专业病理医师进行读片和病理诊断,记录 Gleason 评分结果。

2. MRI 检查

采用 1.5T Philips intera achieva 磁共振成像仪配合心脏阵列线圈,扫描范围包括整个前列腺和精囊。轴面高分辨率 T₂WI: TR 400 ms, TE 12 ms, 层厚 3 mm, 层间距 0 mm, 激励次数 6~8, 视野 200 mm, 矩阵 256×256。DWI: 采用回波平面成像序列(EPI)序列;扫描定位线与轴面 T₂WI 保持一致; b 值: 0, 800 s/mm²; TR 2500 ms, TE 54 ms, 层厚 3 mm, 层间

距 0 mm, 激励次数 8, 视野 200 mm, 矩阵 128×128;扫描完成后自动生成 ADC 图。DCEI: 采用 3D-T₁FFE-WATS 技术, 定位线与轴面 T₂WI 相一致; TR 9.9 ms, TE 5.0 ms, 层厚 4 mm, 层间距 0 mm, 激励次数 1, 视野 400 mm, 矩阵 224×256。在扫描开始的同时使用高压注射器经肘静脉团注对比剂 Gd-DTPA(欧乃影, GE 药业公司), 剂量 0.2 mmol/kg, 流率 2.5 mL/s, 对比剂注入前、后分别注入 21 mL 生理盐水。

3. 图像分析

由两位经过 PI-RADS 应用培训的不同诊断经验医师(Reader 1: 低年资医师, 从事前列腺 MRI 诊断 ≤ 1 年, 前列腺 mp-MRI 阅片数 < 300 例; Reader 2: 高年资医师, 从事前列腺 MRI 诊断 10 年以上, 前列腺 mp-MRI 阅片数 > 3000 例)按照 PI-RADS v1 和 v2 评分标准, 独立观察、盲法分析入组病例的 mp-MRI 图像, 分别对外周带病灶、移行带病灶进行评分。为避免干扰, 同一医师采用两版 PI-RADS 对同一病例进行诊断评分的时间间隔 > 6 周, 记录两位医师的 PI-RADS v1 和 v2 诊断评分结果。

PI-RADS v1 评分标准采用各个序列相加后的方案^[1,3,4], 轴面高分辨率 T₂WI、DWI 和 DCEI 3 个序列单独评分后相加得出总评分, 范围为 3~15 分, 然后再换算成 5 分制(3, 4=1; 5, 6=2; 7~9=3; 10~12=4; 13~15=5)。PI-RADS v2 评分标准则采取 1~5 分制^[2,4,5], 对于前列腺外周带病灶以 DWI 结果为主, 当 DWI 评分为 1、2、4、5 时, 该分即为最终评分。当 DWI 评为 3 分时, 再结合 DCEI 的表现, 若 DCEI 为阳性, 则最终 PI-RADS 评分为 4 分; 若 DCEI 为阴性, 则最终评分仍维持 3 分。移行带病灶以 T₂WI 结果为主, 当 T₂WI 评分为 1、2、4、5 时, 该分即为最终评分。当 T₂WI 评为 3 分时, 应结合 DWI 评分。只有 DWI 评分为 5 分时, 总分才改为 4 分, 否则总分仍为 3 分。

4. 统计学方法

采用 MedCalc 15.0 统计学软件进行加权 Kappa 检验比较不同版本 PI-RADS 情况下两位评分者评分结果间的一致性,将二次加权 $k \leq 0.20$ 定义为一致性较差, k 值介于 $0.21 \sim 0.40$ 为一致性一般水平, k 值介于 $0.41 \sim 0.60$ 为一致性中等水平, k 值介于 $0.61 \sim 0.80$ 则为一致性良好, $k \geq 0.81$ 定义为一致性很好^[6], 通过 Bland-Altman 法画散点图作进一步较为直观的比较^[7]。

结 果

本研究共纳入符合入组标准的前列腺癌 175 例, 其中外周带癌 116 例, 移行带癌 59 例, 年龄 $52 \sim 92$ 岁, 平均 (73.45 ± 6.59) 岁。

两位不同诊断经验的阅片医师采用 PI-RADS v1 对前列腺移行带病灶的评分结果间的一致性仅为一般水平(二次加权 k 值 0.337 , 95% 可信区间 $0.105 \sim 0.568$), 两位医师采用 PI-RADS v2 对移行带病灶的评分结果间的一致性处于中等水平(二次加权 k 值 0.498 , 95% 可信区间 $0.273 \sim 0.723$); 两位医师采用 PI-RADS v1 对外周带病灶的评分结果间的一致性为中等水平(二次加权 k 值 0.417 , 95% 可信区间 $0.252 \sim 0.582$), 两位医师采用 PI-RADS v2 对外周带病灶的评分结果间的一致性为良好水平(二次加权 k 值 0.658 , 95% 可信区间 $0.521 \sim 0.795$); 在不区分外周带及移行带病灶的综合评价中, 两位医师 PI-RADS v1 评分结果一致性为中等水平(二次加权 k 值 0.398 , 95% 的可信区间为 $0.261 \sim 0.535$), 而两位医师 PI-RADS v2 评分结果一致性达到良好水平(二次加权 k 值 0.632 , 95% 的可信区间为 $0.510 \sim 0.755$)。两位医师采用 PI-RADS v1 和 v2 对外周带病灶的诊断评分结果间的一致性均高于移行带病灶(表 1)。

将两位阅片医师分别采用 PI-RADS v1 和 PI-RADS v2 对外周带和移行带病灶评分诊断结果绘制 Bland-Altman 散点图(图 1~3), 结果表明在两位医师评分差值的平均值中, PI-RADS v2 对外周带病灶、移行带病灶评分结果的一致性较 PI-RADS v1 更接近于代表差值为零的线, 提示 v2 评分结果的一致性高于 v1。

讨 论

不同于 PI-RADS v1, PI-RADS v2 采用了更为简

化的评分方式, 针对前列腺分带结构提出了相应的评分权重序列(外周带以 DWI 为主, 移行带以高分辨率 T_2 WI 为主); 弱化了 DCEI 在评分系统中的作用, DCE 序列仅作为备选序列(阳性与阴性强化), 外周带评分为 3 分时才发挥作用; 在高分辨率 T_2 WI 和 DWI 癌灶评分的 4 分与 5 分的评价中引入了病灶大小标准; 在 mp-MRI 的序列构成中剔除了磁共振波谱成像(magnetic resonance spectroscopic imaging, MRSI)。推出 PI-RADS v2 的目的在于建立前列腺 Mp-MR 最简化且可接受的技术和诊断标准, 从而有利于前列腺 MRI 的临床应用与推广^[2,4]。自 PI-RADS v1 和 v2 推出以后, 诸多研究者进行了诊断一致性研究, 其中 Schimmöller 等^[6] 和 Kasel-Seibert 等^[8] 发现不同医师利用 PI-RADSv1 对前列腺癌的评分结果一致性水平要高于对良性病灶的评分, 后者还发现不同阅片医师间对外周带病灶评分结果的一致性要低于移行带病灶的评分; Rosenkrantz 等^[9] 对 PI-RADS v2 在多中心阅片医师间的一致性进行了研究, 结果表明采用 PI-RADS v2 在不同评分者间评分结果的一致性, 外周带较高(k 值最高为 0.619), 移行带较低(k 值最高为 0.529)。本研究不同阅片者采用 PI-RADS v1 和 v2 对外周带癌进行评分诊断时的一致性均高于移行带癌, 与 Rosenkrantz 等的研究结果相符。笔者认为 PI-RADS 本质上是标准化了的主观分级评价方法, 移行带良性增生结节的扩散受限和 DCEI 异常强化表现, 与 PCa 的 DWI 和 DCEI 表现间存在一定程度的重叠, 在此情况下, 诊断经验的高低势必影响到评分结果, 因此导致了无论是采用 PI-RADS v1 还是 v2, 在不同诊断经验的阅片者间评分结果的一致性移行带均低于外周带。

在诸多比较 PI-RADS v1 和 v2 两个版本诊断一致性的研究中, 结果不尽一致, Polanec 等^[10] 的研究结果表明无论 PI-RADS v1 还是 PI-RADS v2 在不同研究者间均具有较佳的一致性(k 值: v1: 0.81 , v2: 0.71)。Greer 等^[11] 对 5 位不同专业领域影像诊断医师 PI-RADS v2 评分结果的一致性进行比较研究, 结果表明在前列腺专科影像诊断医师与非前列腺专科影像诊断医师间具有较高的一致性。Vaché 等^[12] 对不同阅片医师在 PI-RADS v1、PI-RADS v2 和 Likert 量表的评分结果间的一致性进行研究, 结果表明 Likert 量

表 1 两位阅片医师在不同病灶区域的 PI-RADS 一致性结果

标准	TZ 病灶 k 值(95%CI)	PZ 病灶 k 值(95%CI)	所有病灶 k 值(95%CI)
V1	0.337(0.105~0.568)	0.417(0.252~0.582)	0.398(0.261~0.535)
V2	0.498(0.273~0.723)	0.658(0.521~0.795)	0.632(0.510~0.755)

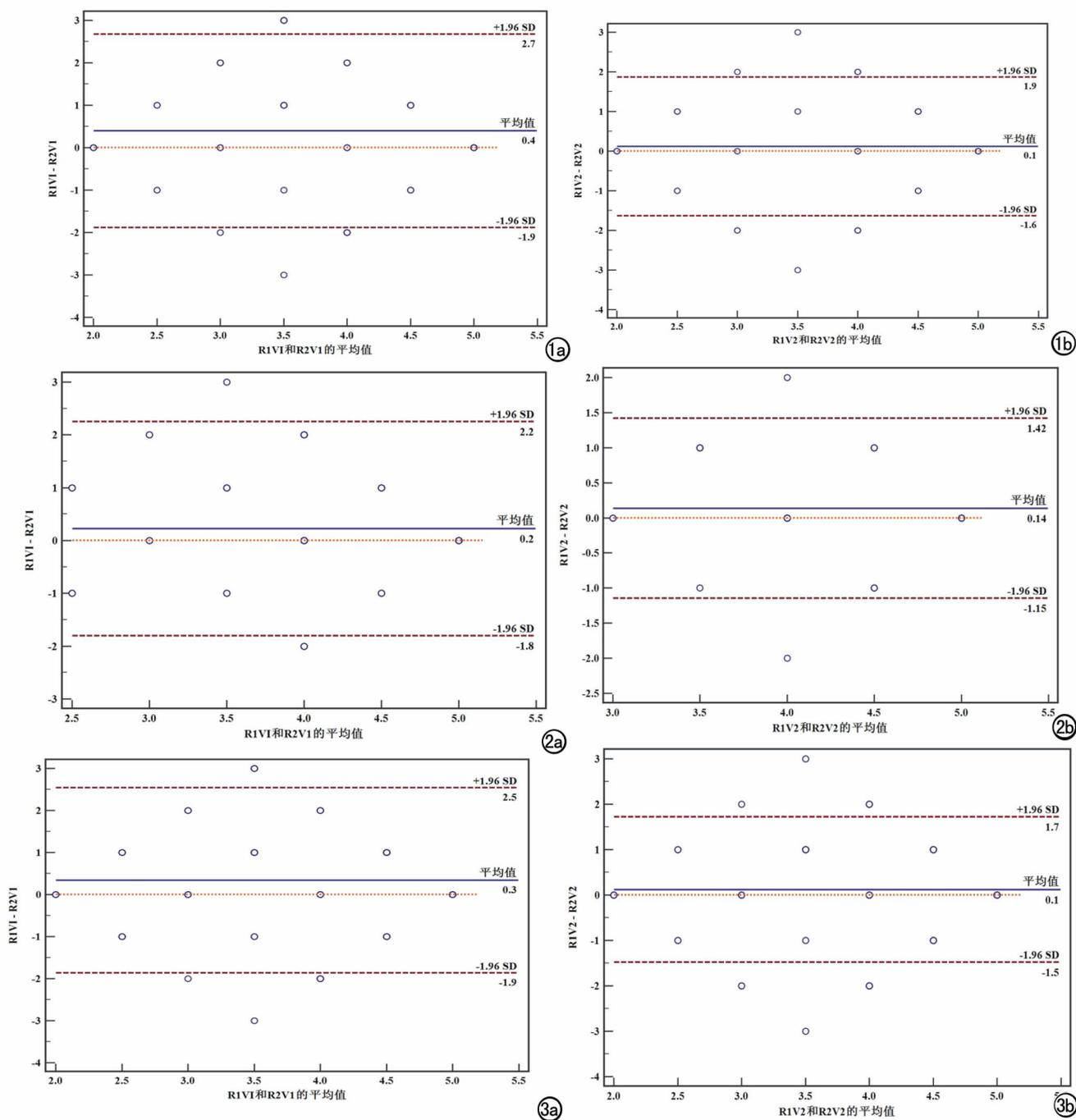


图1 两位阅片医师(R1、R2)分别采用PI-RADS v1和PI-RADS v2对PZ病灶的评分结果绘制Bland-Altman散点图,横坐标为R1、R2评分均值,纵坐标为R1、R2评分差值,横坐标为R1、R2评分差值的平均值,上下的水平虚线是差值的 ± 1.96 倍标准差,代表95%一致性界限的上下限,中间蓝色水平实线代表评分差值的平均值,红色水平虚线代表评分差值的平均值为0,均值线越接近0值,说明两位阅片医师的评分结果间的一致性越高。b所示蓝色水平实线较a更接近于代表差值为零的红色水平虚线,提示对PZ病灶评分中PI-RADS v2较PI-RADS v1在两位诊断经验不同的阅片医师的评分结果间的一致性更高。(R1:Reader 1;R2:Reader 2;V1:PI-RADS v1;V2:PI-RADS v2) 图2 两位阅片医师(R1、R2)分别采用PI-RADS v1和PI-RADS v2对TZ病灶的评分结果绘制Bland-Altman散点图,b所示蓝色水平实线较a更接近于代表差值为零的红色水平虚线,提示对TZ病灶评分中PI-RADS v2较PI-RADS v1在两位诊断经验不同的阅片医师的评分结果间的一致性更高。 图3 两位阅片医师(R1、R2)分别采用PI-RADS v1和PI-RADS v2对所有病灶(PZ+TZ)的评分结果绘制Bland-Altman散点图,b所示蓝色水平实线较a更接近于代表差值为零的红色水平虚线,提示对所有病灶(PZ+TZ)评分中PI-RADS v2较PI-RADS v1在两位诊断经验不同的阅片医师的评分结果间的一致性更高。

表的一致性高于 PI-RADS v1 和 v2。而本研究发现在 1.5T MRI 设备和非直肠内线圈条件下,两位诊断经验相差较大的阅片医师(分别为 ≤ 1 年和 > 10 年)对前列腺病灶应用 PI-RADS v2 评分结果间的一致性要优于 PI-RADS v1,与 Kasel-Seibert 等^[8]的研究结果一致,但即便是表现最好的 PI-RADS v2,两位阅片医师的一致性水平也仅介于中等到良好程度之间,没有出现一致性很好的情况,与 Muller 等^[13]的研究结果类似,提示 PI-RADS v2 的临床应用效能可能会受不同经验医师诊断水平的影响。但上述各研究文献的研究对象和 MRI 设备等的差异,也可能导致研究结果间存在偏差。

总的来说,作为一种偏主观性的评分标准,在 PI-RADS 的临床实际运用中,不同经验阅片医师的诊断评分结果一致性会受到病灶部位和 PI-RADS 版本的影响。与 PI-RADS v1 相比,PI-RADS v2 采用了更为简化的诊断流程,给出了针对前列腺不同区域的权重评分序列,容易被大多数阅片医师,尤其是低年资者所接受,但仍需多中心大样本研究的证实和不断的修正和完善。

本研究的局限性:①仅采用两名诊断经验相差较大的医师来进行一致性分析,缺乏对诊断经验类似水平的不同评分者一致性评估;②入组病例的构成不均衡,移行带癌病例数偏少;③病理结果均为穿刺活检证实。

参考文献:

- [1] Barentsz JO, Richenberg J, Clements R, et al. ESUR prostate MR guidelines 2012[J]. *Eur Radiology*, 2012, 22(4): 746-757.
- [2] Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate imaging-reporting and data system: 2015, Version 2[J]. *Eur Urology*, 2016, 69(1): 16-40.
- [3] 邓明, 王良, 胡道予, 等. 前列腺影像报告和数据系统指南(PI-

RADS)解读及典型病例分析[J]. *放射学实践*, 2013, 28(10): 998-1001.

- [4] 李拔森, 王良. 第二版前列腺影像报告和数据系统(PI-RADS)解读[J]. *中华放射学杂志*, 2015, 49(10): 798-800.
- [5] Feng ZY, Liang W, Min XD, et al. Prostate cancer detection with multiparametric magnetic resonance imaging: prostate imaging reporting and data system version 1 versus version 2[J]. *Chin Med J (Engl)*, 2016, 129(20): 2451-2459.
- [6] Schimmöller L, Quentin M, Arsov C, et al. Inter-reader agreement of the ESUR score for prostate MRI using in-bore MRI-guided biopsies as the reference standard[J]. *Eur Radiol*, 2013, 23(11): 3185-90.
- [7] 陈卉. Bland-Altman 分析在临床测量方法一致性评价中的应用[J]. *中国卫生统计*, 2007, 24(3): 308-309.
- [8] Kasel-Seibert M, Lehmann T, Aschenbach R, et al. Assessment of PI-RADS v2 for the detection of prostate cancer[J]. *Eur J Radiol*, 2016, 85(4): 726-731.
- [9] Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. Interobserver reproducibility of the PI-RADS version 2 lexicon: a multicenter study of six experienced prostate radiologists[J]. *Radiology*, 2016, 280(3): 793-804.
- [10] Polanc S, Helbich TH, Bickel H, et al. Head-to-head comparison of PI-RADS v2 and PI-RADS v1[J]. *Eur J Radiol*, 2016, 85(6): 1125-1131.
- [11] Greer MD, Brown AM, Shih JH, et al. Accuracy and agreement of PIRADSv2 for prostate cancer mpMRI: A multireader study[J]. *J Magn Reson*, 2017, 45(2): 579-585.
- [12] Vaché T, Bratan F, Mègelechevallier F, et al. Characterization of prostate lesions as benign or malignant at multiparametric MR imaging: comparison of three scoring systems in patients treated with radical prostatectomy[J]. *Radiology*, 2014, 272(2): 446-455.
- [13] Muller BG, Shih JH, Sankineni S, et al. Prostate cancer: interobserver agreement and accuracy with the revised prostate imaging reporting and data system at multiparametric MR imaging[J]. *Radiology*, 2015, 277(3): 741-750.

(收稿日期:2018-08-02 修回日期:2019-01-13)