

# Concordancia Interobservador para la Validación del PI-RADS v2

## *Interobserver Agreement for the PI-RADS v2 Validation*

Andrés Labra W.<sup>1</sup> Claudio Silva Fuente-Alba<sup>1</sup> Giancarlo Schiappacasse F.<sup>1</sup> Daniela Barahona Z.<sup>1</sup>  
Velimir Skoknic B.<sup>2</sup>

<sup>1</sup> Radiólogo de la Facultad de Medicina Clínica Alemana de Santiago, Universidad del Desarrollo, Vitacura, Chile

<sup>2</sup> Residente de Radiología Facultad de Medicina Clínica Alemana de Santiago, Universidad del Desarrollo, Vitacura, Chile

**Address for correspondence** Andrés Labra W., MD, Radiólogo Facultad de Medicina, Clínica Alemana de Santiago, Universidad del Desarrollo, Av. Vitacura 5951, Vitacura, Chile (e-mail: alabra@alemana.cl).

Rev Argent Radiol 2019;83:49–55.

### Resumen

**Objetivo** Evaluar la variabilidad interobservador en el uso de la versión 2.0 del PI-RADS (PI-RADS v2), en lectores experimentados y no experimentados.

**Materiales y Métodos** Estudio retrospectivo de análisis de concordancia de lectores. Entre enero de 2015 y diciembre de 2016, 1.656 sujetos fueron estudiados mediante resonancia magnética multiparamétrica (RMmp) de próstata en nuestra institución. Se estimó la distribución porcentual del reporte en categoría PI-RADS v2, y con esa información, se realizó una selección de 150 casos con esquema de aleatorización estratificada a las distribuciones porcentuales de cada categoría. Dichos casos fueron anonimizados, presentados a tres lectores con cinco, cuatro y dos años de experiencia en lectura de RMmp además de tener más de un año de experiencia en el uso del PI-RADS v2 siendo leídos en forma individual. Los datos resultantes fueron analizados en forma independiente por un cuarto investigador.

**Resultados** El valor de kappa ponderado para los observadores fue de 0,69 (IC 95: 0,64 a 0,75). La mayor concordancia correspondió a los lectores de mayor experiencia, donde alcanza un valor de 0,72 (IC 95%: 0,69 a 0,76). La concordancia entre los valores PI-RADS que determinan seguimiento o bien una intervención de acuerdo a elementos clínicos (1–2–3) y conducta activa (4–5) correspondió a 0,70 (IC 95%: 0,59 a 0,78).

**Discusión** Se logró demostrar un acuerdo sustancial entre radiólogos utilizando el PI-RADS v2 para la detección en RMmp de lesiones sospechosas, mayor entre los dos lectores más experimentados. Sin embargo, la comparación del lector de menor experiencia con los de mayor experiencia también presentó una importante concordancia. Los valores de concordancia entre observadores para PI-RADS  $\geq 4$  fueron similares a los reportados en la literatura.

### Palabras clave

- ▶ próstata
- ▶ resonancia magnética
- ▶ variación interobservador
- ▶ PI-RADS

Andrés Labra W.'s ORCID is <https://orcid.org/0000-0002-7060-6361>.

Velimir Skoknic B.'s ORCID is <https://orcid.org/0000-0002-4958-2131>.

received  
September 10, 2018  
accepted  
May 8, 2019

DOI <https://doi.org/10.1055/s-0039-1692206>.  
ISSN 1852-9992.

Copyright © 2019, Sociedad Argentina de Radiología. Publicado por Thieme Revinter Publicações Ltda., Rio de Janeiro, Brazil. Todos los derechos reservados.

License terms



## Abstract

### Keywords

- ▶ prostate
- ▶ magnetic resonance imaging
- ▶ interobserver variation
- ▶ PI-RADS

**Conclusiones** El PI-RADS v2 ha demostrado en nuestro centro, con radiólogos dedicados a imágenes de abdomen y estudios de próstata, un alto nivel de acuerdo en la interpretación de la RMmp de próstata, encontrándose a tono con lo reportado en la literatura.

**Purpose** To evaluate the interobserver variability in the use of the PI-RADS 2.0 version, in experienced and non-experienced readers.

**Material and Methods** Retrospective study of readers' concordance analysis. Between January 2015 and December 2016, 1656 subjects were studied through multiparametric MRI (RMmp) of prostate in our institution. The percentage distribution of the report was estimated in each PI-RADS category, and a selection of 150 cases with a stratified randomization scheme was made to the percentage distributions of every category. These cases were anonymized, presented to three readers with 5, 4 and 2 years' experience in reading RMmp, and more than one-year experience with PI-RADS v2, and were read individually. The resulting data were analyzed independently by a fourth investigator.

**Results** The weighted kappa value for the observers was 0.69 (IC 95: 0.64 to 0.75). The highest agreement corresponded to the two most experienced readers, where it reached a value of 0.72 (95% CI: 0.69 to 0.76). The concordance between the PI-RADS values that determine follow-up (1–2–3) and active behavior (4–5) corresponded to 0.70 (95% CI: 0.59 to 0.78).

**Discussion** It was possible to demonstrate a substantial agreement between radiologists using the PI-RADS v2 for the detection in RMmp of suspicious lesions, greater among the two most experienced readers. However, the comparison of the less experienced reader with those of greater experience also presented an important concordance. The interobserver concordance values for PI-RADS  $\geq 4$  were like those reported in the literature.

**Conclusions** The PI-RADS v2 has demonstrated, in our center, with radiologists dedicated to abdominal images and prostate studies, a high level of agreement in the interpretation of prostate MRmp.

## Introducción

La resonancia magnética multiparamétrica (RMmp) y las biopsias dirigidas han mejorado la posibilidad de detección de cánceres de próstata clínicamente significativos, reduciendo así el diagnóstico de cánceres no significativos o de bajo riesgo.<sup>1</sup> Sin embargo, para que la RMmp tenga mayor impacto, requiere un alto nivel de experiencia en su interpretación, lo que ha determinado una significativa variabilidad interobservador. Eso se puede deber, en parte, a criterios no estandarizados para el diagnóstico de las anomalías que se describen en la RMmp.<sup>2</sup>

En el año 2015, un esfuerzo conjunto entre el Colegio Americano de Radiología (ACR), la Sociedad Europea de Radiología Urogenital (ESUR) y la Fundación AdMeTech, publicó la segunda versión de "Prostate Imaging Report and Data System" (PI-RADS v2), cuyas guías están destinadas a mejorar la estandarización en la interpretación de la RMmp de próstata.<sup>3</sup> La versión 2.0 del PI-RADS (PI-RADS v2) entrega información para la adquisición e interpretación de la RMmp de próstata y propone un sistema simplificado en escala y score de cinco puntos, que indican la probabilidad de existencia de un cáncer clínicamente significativo. Ese sistema se creó

para mejorar la precisión del acuerdo interobservador del PI-RADS v1, que presentaba significativas limitaciones por variabilidades en la interpretación, con una escala poco clara con suma de diferentes scores (escala de 3 a 15) y por lo tanto, poca claridad en identificar cánceres clínicamente significativos.<sup>4</sup>

El PI-RADS v2 incluye numerosos cambios que se describen a continuación:

- (a) Se introdujo el concepto de secuencia dominante dependiendo de la localización de la lesión. En zona periférica la secuencia dominante es la difusión (DWI) con su mapa ADC; en cambio, en la zona de transición la secuencia dominante es el T2.
- (b) La perfusión (DCE) se debe de reportar como positiva cuando existe una impregnación focal precoz y negativa cuando no existe ese tipo de impregnación o es heterogénea o difusa. No es necesario el análisis de curvas de perfusión como estaba descrito en la versión original del PI-RADS.
- (c) Cuando la perfusión es positiva, el score final se puede incrementar en un punto, pero solo si representa una diferencia clínica significativa, como es el caso de incrementar un PI-RADS 3 a score 4.

(d) Un *score* final con escala de 1 a 5 se asigna de acuerdo con las reglas revisadas en el PI-RADS v2.

El objetivo de este estudio es evaluar la variabilidad interobservador en el uso de la versión 2.0 del PI-RADS. Una adecuada implementación de ese sistema puede ayudar a los médicos clínicos a estimar el riesgo de cáncer de próstata clínicamente significativo y, por lo tanto, tomar una conducta adecuada para la realización de biopsias, ya sean sistemáticas o dirigidas, a una lesión significativa por RMmp.<sup>5</sup>

## Materiales y Métodos

Estudio retrospectivo de análisis de concordancia de lectores, aprobado por el Comité de Ética Institucional que, atendidas las características del estudio, autorizó una exención del consentimiento informado.

Entre enero de 2015 y diciembre de 2016, 1.656 sujetos fueron estudiados mediante RMmp de próstata en nuestra institución. Los estudios fueron realizados en resonador magnético 3 Tesla (Magnetom Skyra, Siemens Healthcare, Erlangen, Germany) plataforma versión Numaris/4. Se utilizó bobina de superficie *phase array* de pelvis de 30 canales, con administración de buscapina 10 mg ev, inmediatamente previo a la realización del estudio. El protocolo de estudio incluyó secuencias ponderadas en T2 TSE multiplanares sagital, coronal y axial (TR/TE: 4780/90, FOV 18 cm, 3/0 mm, 320/272), T1 TSE axial, difusión (3/0 mm, b0, 50, 500, 1000 y 1600), mapa ADC y estudio de perfusión (resolución temporal 7 seg, 3 mm).

Dentro de esos sujetos, se estimó la distribución porcentual del reporte en categoría PI-RADS v2. Con esa información, se realizó una selección de 150 casos con esquema de aleatorización estratificada a las distribuciones porcentuales de categorías de PI-RADS durante los años 2015 y 2016. No se contó con ningún criterio de exclusión.

Dichos casos fueron anonimizados y presentados a tres lectores que tenían cinco, cuatro y dos años de experiencia en la lectura de RMmp, además de tener más de un año de experiencia en el uso del PI-RADS v2. Cada lector dispuso de todas las imágenes adquiridas en cada estudio. Los casos fueron aleatorizados en secuencia, de forma que cada lector disponía de los mismos casos, pero presentados en un orden secuencial diferente. Se utilizaron todas las secuencias por cada lector para evaluar cada caso e interpretar el PI-RADS.

Fueron leídos en forma individual y en cinco sesiones independientes de 30 casos en plataforma Impax 6.5.2.657 (Agfa®) y los datos resultantes fueron analizados en forma independiente a los lectores por un cuarto investigador. No disponíamos, al momento de la lectura individual, de los valores de antígeno prostático específico (APE) ni de los datos de la indicación clínica por la que se solicitó el estudio.

El acuerdo entre los lectores fue analizado usando la estadística de kappa ponderada modificada desde Cohen, considerando las categorías de acuerdo en base a las recomendaciones de Landis y Koch (0.00 Pobre; 0.01–0.20 Leve; 0.21–0.40 Aceptable; 0.41–0.60 Moderada; 0.61–0.80

Importante; 0.81–1.00 Casi perfecta). Las características cualitativas son descritas por sus distribuciones de porcentaje.

Para el análisis estadístico se utilizó el *software* Stata (Version 14.0, StataCorp, College Station, Texas) Se consideró diferencia estadísticamente significativa con  $p < 0.05$  y se calcularon intervalos de confianza de 95% según correspondía.

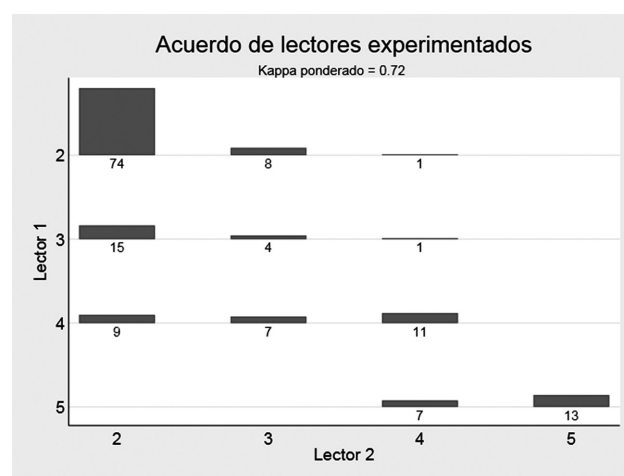
## Resultados

Los sujetos tuvieron una mediana de edad de 60 años (Rango Inter Cuartil (RIC): 55–67) y la distribución de los 150 casos considerados incluyeron 79 PI-RADS 2, 28 PI-RADS 3, 24 PI-RADS 4 y 19 PI-RADS 5 (estratificados desde la distribución porcentual de la población total).

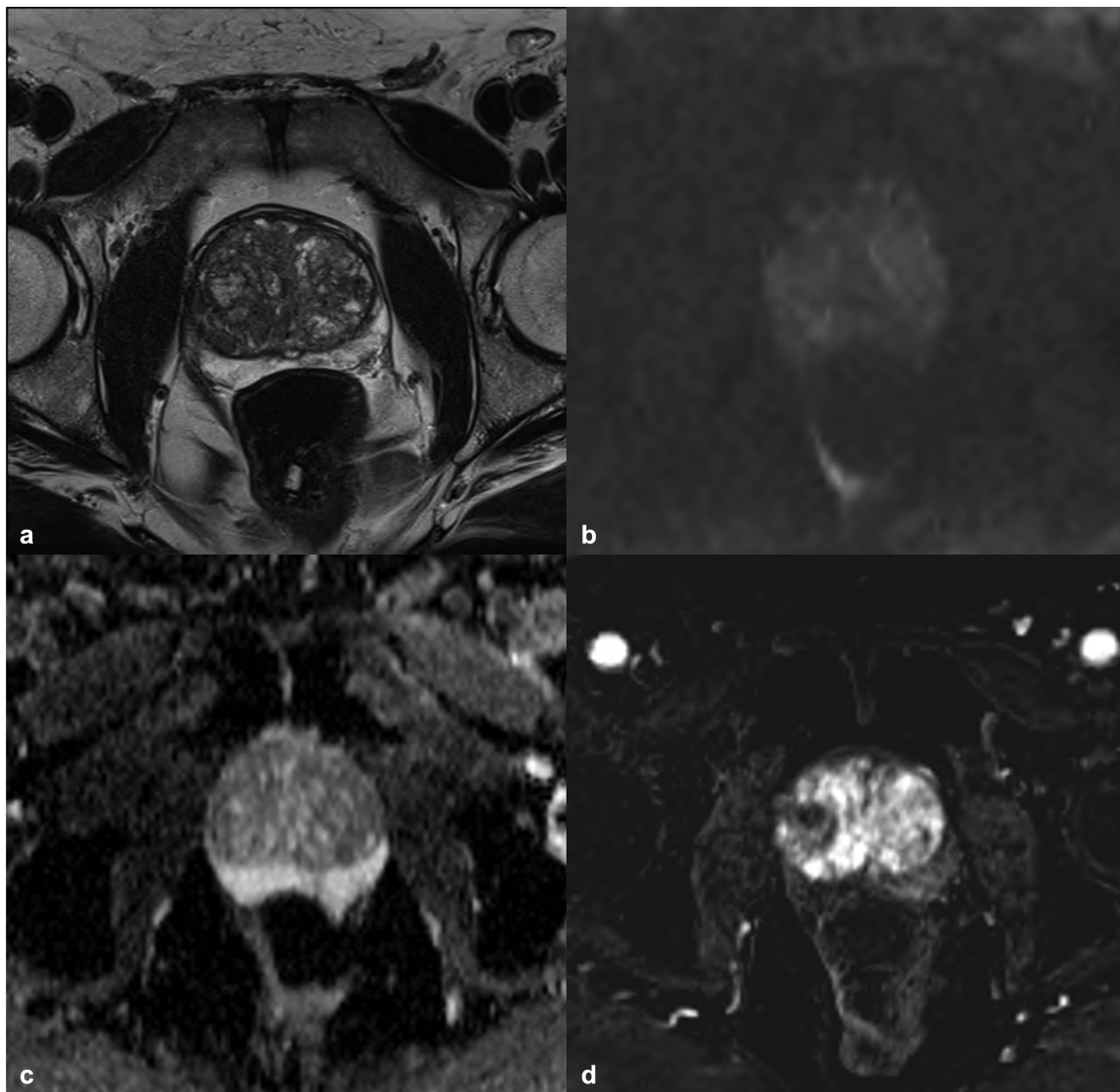
Todos los estudios fueron considerados de características diagnósticas por los tres lectores.

El valor de kappa ponderado para los tres observadores fue de 0.69 (IC 95: 0.64 a 0.75). Destaca que la mayor concordancia correspondió a los dos lectores de mayor experiencia donde alcanza un valor de 0.72 (IC 95%: 0.69 a 0.76) (►Figura 1). La concordancia entre el lector de menor experiencia con los dos lectores más expertos en forma individual varió entre 0.680 con lector 1 (IC 95% 0.61 a 0.71) y 0.675 (IC 95% 0.60 a 0.70) con el lector 2. Todos esos valores caen en la categoría de importante acuerdo (según Landis y Koch). Las ►Figuras 2 y 3 muestran casos representativos de concordancia entre los tres lectores en casos de PI-RADS 2 y PI-RADS 5 respectivamente.

La concordancia entre los valores PI-RADS que determinan el seguimiento (1–2–3) y la conducta activa (4–5), correspondió a 0.70 (IC 95%: 0.59 a 0.78) (►Figura 4). En ese escenario, la concordancia entre los lectores de mayor experiencia fue de 0.71 (IC 95%: 0.58 a 0.84) (►Figura 4) y la concordancia entre el lector de menor experiencia con los dos lectores más expertos en forma individual, varió entre 0.697 con lector 1 (IC 95% 0.57 a 0.82) y 0.679 (IC 95% 0.55 a 0.81) con el lector 2, nuevamente caen en la categoría de importante acuerdo. La ►Figura 5 muestra un caso



**Fig. 1** Gráfico que muestra la distribución de los resultados de las interpretaciones de los lectores experimentados (cinco y cuatro años respectivamente), con un kappa ponderado de 0,72 (concordancia importante).



**Fig. 2** Paciente de 56 años. APE 5,7 ng/ml (a) T2FSE WI, (b, c) DWI y mapa ADC, (d) perfusión (DCE). Signos de hiperplasia de la zona de transición con nódulos adenomatosos. PI-RADS 2, existiendo acuerdo entre los tres lectores.

representativo de discordancia entre un lector de menor experiencia con los dos lectores de mayor experiencia.

## Discusión

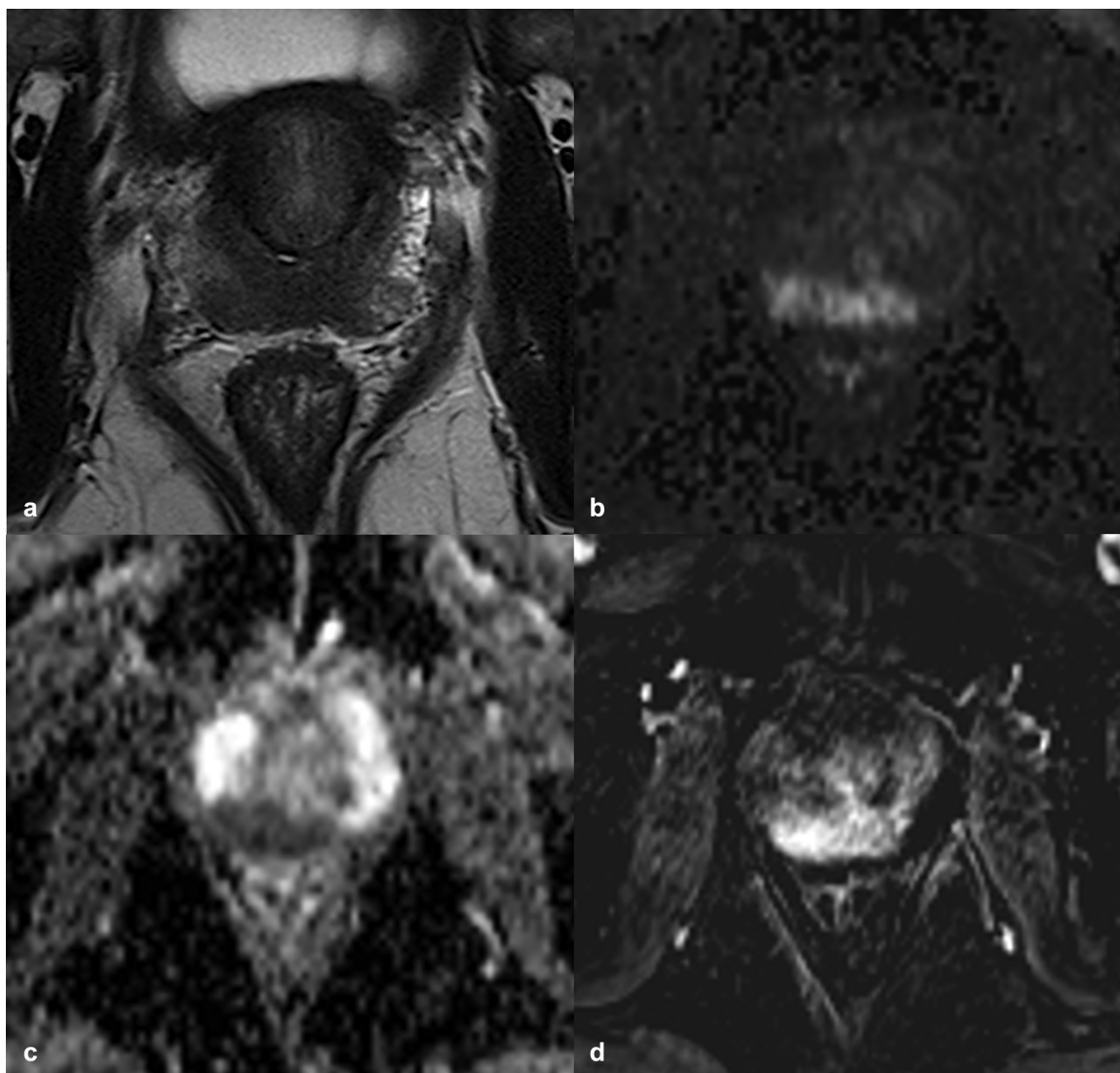
En este estudio se logró demostrar un acuerdo importante entre tres radiólogos utilizando el PI-RADSV2 para la detección en RMmp de lesiones sospechosas, siendo la concordancia mayor entre los dos lectores más experimentados. Sin embargo, la comparación del lector de menor experiencia con los de mayor experiencia también presentó una importante concordancia, ya sea con cualquiera de los otros dos lectores o con el conjunto de ellos. Esos niveles de acuerdo son similares a los descritos en la literatura, siendo muy importante el hecho de que en este estudio, se incluyó a un lector de menor experiencia, lo que aporta más validez a la aplicación clínica del PI-RADS.

Los valores de concordancia entre observadores para PI-RADS  $\geq 4$  fueron similares a los reportados en la literatura,

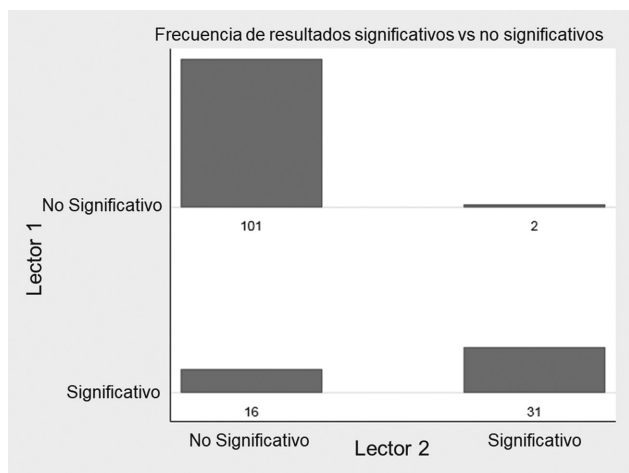
aportando confiabilidad de este método de estudio en la detección de cánceres significativos.<sup>6-9</sup>

Los resultados de este estudio, indican que lectores con diferentes niveles de experiencia, pueden utilizar el PI-RADS v2 para la detección de lesiones sospechosas clínicamente significativas con alto acuerdo entre los observadores. Eso es de gran utilidad, ya que así los médicos clínicos cuentan con más información para tomar una actitud adecuada y confiable a partir de los hallazgos descritos en las RMmp, pudiendo definir si indicar o no a los pacientes biopsia transrectal sistemática dirigida a la lesión sospechosa en forma cognitiva o bien por fusión RM ultrasonido (RM/US).<sup>3,5,8</sup>

A diferencia de los trabajos de Rosenkrantz y col.,<sup>6</sup> y de Girometti y col.,<sup>9</sup> en los que se realizó un análisis de concordancia histopatológica junto a la concordancia interobservador, en este estudio, no se realizó esa correlación histopatológica, así como tampoco un seguimiento a largo plazo de la evolución de los pacientes, lo que claramente es una limitante de nuestro estudio y entregaría valiosa



**Fig. 3** Paciente de 73 años. PSA 7,8 ng/ml (a) T2 FSE. Lesión hipointensa de 17 mm en Zpposteromedial del LPD, con extensión al lado contralateral (b). Restricción en secuencia de DWI y baja señal en mapa ADC (c), con significativa impregnación en estudio de perfusión (d). Biopsia dirigida por fusión RM/US demuestra adenocarcinoma Gleason 4 + 5 en zona periférica. En ese caso, hubo acuerdo entre todos los lectores que informaron PI-RADS 5.

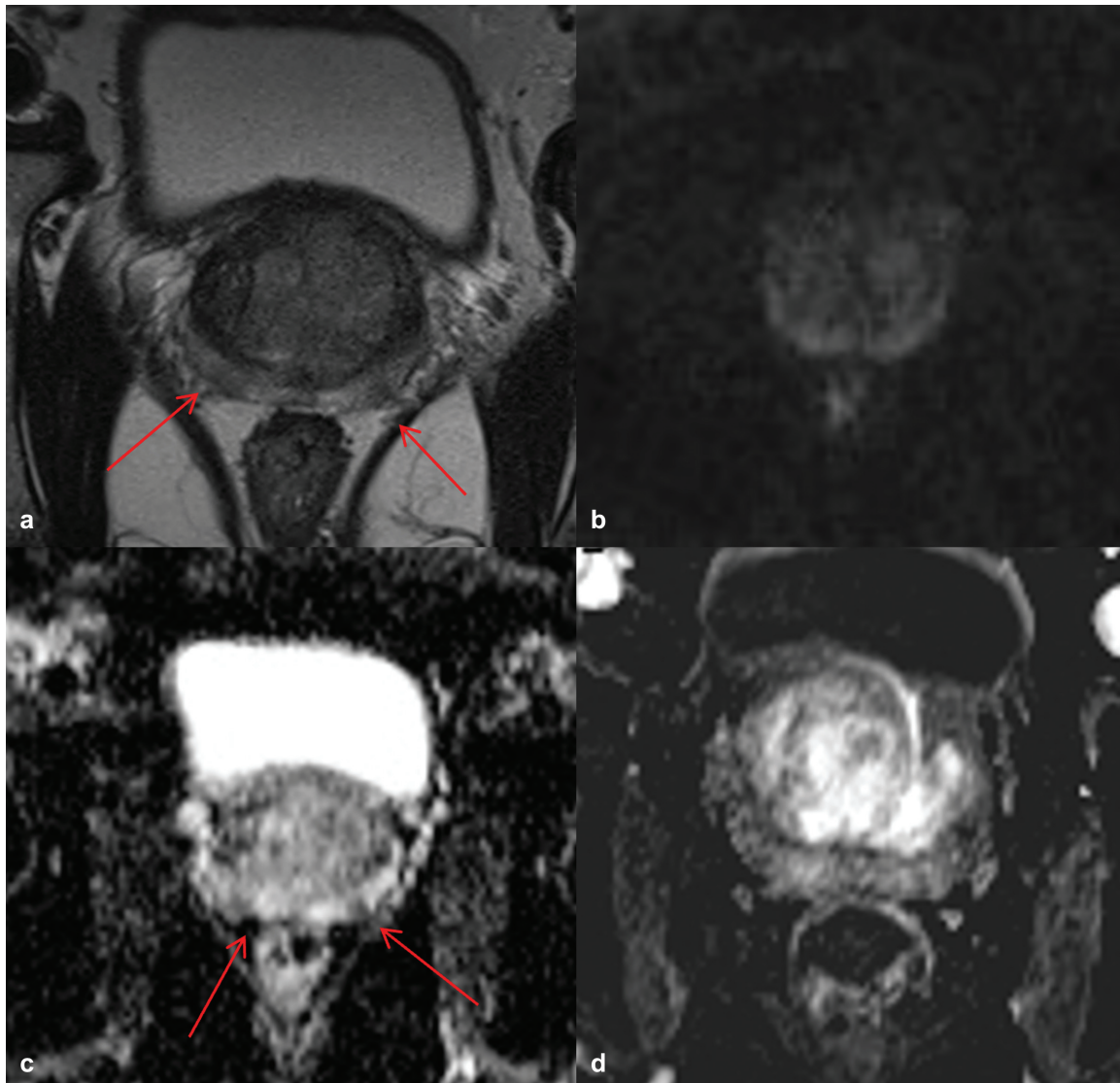


**Fig. 4** Diagrama que muestra la distribución y concordancia de los resultados de las interpretaciones de los lectores experimentados según si los resultados son no significativos (PI-RADS 1–2–3) o significativos (PI-RADS 4–5).

información con respecto a la concordancia interobservador en el contexto de validez diagnóstica. Otro aspecto importante es que se realizó en un mismo centro con un solo equipo y protocolo, por lo que no se puede descartar una mayor variabilidad interobservador en caso de comparar diferentes instituciones y protocolos. Estudios futuros son necesarios para validar el PI-RADS v2 en todo el mundo, especialmente de tipo prospectivo con una distribución de riesgo mayor.

### Conclusión

En conclusión, la versión 2.0 del PI-RADS ha demostrado, en nuestro centro, con radiólogos dedicados a imágenes de abdomen y estudios de próstata, un alto nivel de acuerdo en la interpretación de RMmp de próstata. El PI-RADS v2 es un paso más en la estandarización de la interpretación de RMmp de próstata con lectores de diferentes niveles de experiencia.



**Fig. 5** Paciente de 60 años. PSA 5,1 ng/ml (a) T2 FSE. Pequeñas lesiones hipointensas de 5 mm en zona periférica bilateral. (b) Mínima representación en secuencia de DWI y en mapa ADC (c), sin impregnación focal en estudio de perfusión (d). Biopsia sistemática demuestra adenocarcinoma Gleason 3 + 3 en zona periférica de ambos sextantes. En ese caso, hubo acuerdo entre lectores de mayor experiencia, que informaron PI-RADS 3, y desacuerdo con lector de menor experiencia, que informó PI-RADS 2.

#### Responsabilidades Éticas

**Protección de personas y animales.** Los autores declaran que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

**Confidencialidad de los datos.** Los autores declaran que han seguido los protocolos de su centro de trabajo sobre la publicación de datos de pacientes.

**Derecho a la privacidad y consentimiento informado.** Los autores declaran que en este artículo no aparecen datos de pacientes.

#### Conflicto de Intereses

Los autores declaran no tener ningún conflicto de interés.

#### Bibliografía

- 1 Park SY, Jung DC, Oh YT, et al. Prostate Cancer: PI-RADS Version 2 Helps Preoperatively Predict Clinically Significant Cancers. *Radiology* 2016;280(01):108-116
- 2 Dickinson L, Ahmed HU, Allen C, et al. Scoring systems used for the interpretation and reporting of multiparametric MRI for prostate cancer detection, localization, and characterization: could standardization lead to improved utilization of imaging within the diagnostic pathway? *J Magn Reson Imaging* 2013;37(01):48-58
- 3 Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. *Eur Urol* 2016;69(01):16-40
- 4 Barentsz JO, Richenberg J, Clements R, et al; European Society of Urogenital Radiology. ESUR prostate MR guidelines 2012. *Eur Radiol* 2012;22(04):746-757
- 5 Zhao C, Gao G, Fang D, et al. The efficiency of multiparametric magnetic resonance imaging (mpMRI) using PI-RADS Version 2 in the diagnosis of clinically significant prostate cancer. *Clin Imaging* 2016;40(05):885-888
- 6 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. *Radiology* 2016; 280(03):793-804
- 7 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. *Radiology* 2015;277(03):741–750
- 8 Chen F, Cen S, Palmer S. Application of Prostate Imaging Reporting and Data System Version 2 (PI-RADS v2): Interobserver Agreement and Positive Predictive Value for Localization of Intermediate- and High-Grade Prostate Cancers on Multiparametric Magnetic Resonance Imaging. *Acad Radiol* 2017;24(09):1101–1106
- 9 Girometti R, Giannarini G, Greco F, et al. Interreader agreement of PI-RADS v. 2 in assessing prostate cancer with multiparametric MRI: A study using whole-mount histology as the standard of reference. *J Magn Reson Imaging* 2019;49(02):546–555