



Revista argentina de radiología

ISSN: 1852-9992

ISSN: 0048-7619

Sociedad Argentina de Radiología (SAR) y Federación Argentina de Diagnóstico por Imágenes y Terapia Radiante (FAARDIT)

Múnera-Orozco, Valentina; Isaza-Zapata, Sebastián; Escobar, Luz A.; Peláez-Arroyave, Sebastián; Castrillón, Germán A.

Concordancia inter- e intraobservador del puntaje PI-RADS 2.1 del Colegio Americano de Radiología

Revista argentina de radiología, vol. 88, núm. 1, 2024, Enero-Marzo, pp. 3-10

Sociedad Argentina de Radiología (SAR) y Federación Argentina de Diagnóstico por Imágenes y Terapia Radiante (FAARDIT)

DOI: <https://doi.org/10.24875/RAR.23000003>

Disponible en: <https://www.redalyc.org/articulo.oa?id=382577544007>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org



Sistema de Información Científica Redalyc

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Concordancia inter- e intraobservador del puntaje PI-RADS 2.1 del Colegio Americano de Radiología

Intraobserver and interobserver agreement of the American College of Radiology PI-RADS 2.1 score

Valentina Múnera-Orozco*, Sebastián Isaza-Zapata, Luz A. Escobar,
Sebastián Peláez-Arroyave, Germán A. Castrillón

Centro de Imágenes Diagnósticas, Centro de Diagnóstico Médico, Medellín, Colombia

Resumen

Antecedentes: La resonancia magnética (RM) de próstata es uno de los métodos diagnósticos para la identificación del carcinoma de próstata. La escala PI-RADS (Prostate Imaging and Reporting Data System) es el sistema usado para la interpretación de estas imágenes. Es importante, para su reproducibilidad, la estandarización y la evaluación de dicha escala. **Objetivo:** Determinar la concordancia inter- e intraobservador de la versión 2.1 del PI-RADS. **Material y métodos:** Estudio observacional retrospectivo, evaluando 129 RM de pacientes con sospecha de cáncer de próstata por tres radiólogos con diferentes años de experiencia y en dos momentos del tiempo, usando el puntaje PI-RADS 2.1. Se evaluó la concordancia intra- e interobservador. **Resultados:** La concordancia interobservador fue sustancial ($\kappa > 0,6$) en todos los observadores, siendo la categoría 5 la de mayor acuerdo interobservador. Se observó una alta reproducibilidad intraobservador, con la mayor kappa siendo de 0,856. Cuando se realizó el análisis según años de experiencia de los radiólogos, la concordancia interobservador fue significativa en todos los casos. **Conclusiones:** El sistema de clasificación PI-RADS 2.1 es reproducible para las diferentes categorías y aumenta la concordancia cuando se trata de lesiones con mayor probabilidad de cáncer clínicamente significativo.

Palabras clave: Próstata. Detección temprana de cáncer. Resonancia magnética. Neoplasia de próstata.

Abstract

Background: Magnetic Resonance Imaging (MRI) of the prostate is a key diagnostic tool for identifying prostate carcinoma. The Prostate Imaging-Reporting and Data System (PI-RADS) scale is the standard system for interpreting these images. Standardizing and evaluating this scale is crucial for ensuring consistent and reproducible results. **Objective:** This study aims to assess both the interobserver and intraobserver agreement of the PI-RADS version 2.1. **Material and methods:** In this retrospective observational study, 129 prostate MRI scans from patients with suspected prostate cancer were evaluated. Three radiologists, each with different levels of experience, analyzed these scans at two separate times using the PI-RADS 2.1 scoring system. Both intraobserver and interobserver agreements were measured. **Results:** The study found substantial interobserver agreement ($\kappa > 0.6$) across all categories, with category 5 showing the highest level of agreement. Intraobserver reproducibility was also high, with the highest kappa value reaching 0.856. Further analysis based on the radiologists' years of experience revealed significant interobserver agreement in all instances. **Conclusions:** The PI-RADS 2.1 classification system demonstrates high reproducibility across different categories, particularly for lesions more likely to be clinically significant cancers. This underscores its reliability in varied diagnostic scenarios.

Keywords: Prostate. Early detection of cancer. Magnetic resonance imaging. Prostatic neoplasms.

*Correspondencia:

Valentina Múnera-Orozco

E-mail: valentina.muneraorozco@gmail.com

Fecha de recepción: 12-01-2023

Fecha de aceptación: 11-11-2023

DOI: 10.24875/RAR.23000003

Disponible en internet: 04-03-2024

Rev Argent Radiol. 2024;88(1):3-10

www.revistarar.com

1852-9992 / © 2023 Sociedad Argentina de Radiología (SAR) y Federación Argentina de Asociaciones de Radiología, Diagnóstico por Imágenes y Terapia Radiante (FAARDIT). Publicado por Permanyer. Éste es un artículo open access bajo la licencia CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

El cáncer de próstata es actualmente el segundo cáncer más común en los hombres en todo el mundo. En Colombia es el cáncer de mayor prevalencia y el segundo en mortalidad en la población masculina¹.

En la evaluación del paciente con sospecha de cáncer de próstata se utiliza la biopsia ecodirigida transrectal como estudio de primera línea en aquellos que presentan aumento del valor del antígeno específico de la próstata (PSA) o alteración del tacto rectal². La biopsia ecodirigida transrectal de la próstata es un método con limitaciones, con un rendimiento menor que el esperado, en la detección de cáncer. En el estudio realizado por Serefoglu et al.³, la tasa de detección de cáncer de próstata por biopsia de 12 cilindros tuvo sensibilidades tan bajas como del 33,3%, y se demostró que las biopsias repetidas pueden ser persistentemente negativas a pesar de que exista una lesión tumoral prostática.

En los años recientes, la estandarización de la resonancia magnética (RM) de la próstata ha permitido una mejoría en la localización y la determinación del cáncer, y es un complemento a la biopsia guiada por ecografía para la detección de carcinoma prostático⁴. Al incorporar la RM de próstata como test inicial, antes de la biopsia prostática ecodirigida, es posible disminuir el número de biopsias innecesarias, mejorar la detección de cáncer prostático, disminuir el costo y aumentar la efectividad en el tratamiento⁵.

El sistema de puntuación PI-RADS (*Prostate Imaging Reporting and Data System*) se crea para la valoración de las imágenes en pacientes con sospecha de cáncer de próstata. Este sistema de clasificación se basa en los hallazgos anatómicos y funcionales (visualizados en las secuencias potenciadas en T2, difusión y T1 dinámico poscontraste) para identificar lesiones con alta probabilidad de ser un cáncer que amerite tratamiento⁶. Tiene una sensibilidad del 89% y una especificidad del 73% para la detección de lesiones tumorales prostáticas⁷. En el estudio PROMIS, los autores demostraron que la RM es más sensible (93%; intervalo de confianza del 95% [IC95%]: 88-96) que la biopsia guiada por ecografía (48%; IC95%:42-55; $p < 0,0001$) en el diagnóstico de cáncer clínicamente significativo, con menos eventos adversos para los pacientes, al tratarse de un examen no invasivo⁵.

Dado el rol fundamental que adquiere la RM en el diagnóstico del cáncer de próstata clínicamente significativo (definido como un puntaje de Gleason ≥ 7 en la patología, un volumen tumoral $> 0,5$ ml o extensión extraprostática)⁵, existe la necesidad de una

estandarización de su interpretación. Por lo anterior, la última versión del PI-RADS (versión 2.1) surge para unificar y facilitar la interpretación de las imágenes de RM⁶.

La finalidad de la implementación del PI-RADS en la lectura de las imágenes de RM, en pacientes con sospecha de cáncer de próstata, es permitir la unificación de los términos y que, por medio de la lectura del radiólogo, el clínico pueda hacer mejores estimaciones del riesgo de cáncer de próstata en el paciente⁸. Por ello, es importante evaluar la reproducibilidad de esta escala de clasificación. Algunos estudios previos ya han valorado la concordancia intra- e interobservador, encontrando una reproducibilidad aceptable con valores kappa de 0,43⁸, 0,552⁹ y 0,69¹⁰.

La RM está desempeñando un papel importante en la detección del cáncer de próstata, por lo que resulta necesario evaluar la concordancia inter- e intraobservador de la aplicación de la última versión del sistema de puntuación PI-RADS (v. 2.1) y validar su uso dentro del algoritmo diagnóstico de los pacientes con sospecha de cáncer de próstata.

El propósito de este estudio fue determinar dicha concordancia en tres radiólogos con diferentes años de experiencia.

Método

Protocolo de imagen

Todos los estudios fueron realizados en un resonador magnético marca Siemens Magnetom Skyra de 3.0 Tesla. Las imágenes se adquirieron utilizando bobina corporal. Todos los protocolos de imagen incluyeron secuencia potenciada en T2 TSE (*Turbo Spin Echo*) en los tres planos de imagen (axial, coronal y sagital) y secuencia DWI (*Diffusion-Weighted Imaging*). Se realizaron imágenes potenciadas en T1 con supresión grasa, antes y después de la administración de medio de contraste basado en gadolinio (gadobutrol) a dosis de 0,1 mmol/kg con velocidad de inyección de 3-3,5 ml/s, seguido de 30-40 ml de solución salina fisiológica. Se usó enema rectal dos horas antes del estudio y se administraron 20 mg de N-butil-bromuro de hioscina por vía intravenosa, diez minutos antes de iniciar la adquisición de las imágenes.

Las adquisiciones se realizaron con los factores técnicos descritos en la [tabla 1](#).

Técnica e instrumento de recolección

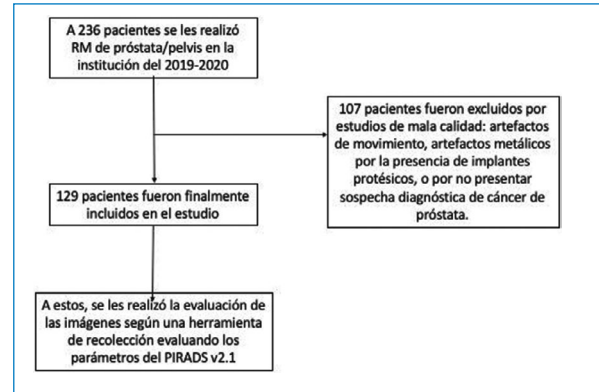
Se llevó a cabo la búsqueda de los estudios de RM de próstata realizados entre los años 2019 y 2020 en

Tabla 1. Factores técnicos en la adquisición de las RM de próstata

Factores técnicos	3T
T2 axial, TSE	FOV de 220 mm TR: 4400 TE: 83 Nex: 2 Matrix: $0,3 \times 0,3 \times 3,5$ mm Espesor de corte: 3,5 mm Cortes: 28
T2, coronal, TSE	FOV de 190 mm TR: 6000 TE: 78 Nex: 1 Matrix: $0,6 \times 0,6 \times 4$ mm Espesor de corte: 4 mm Cortes: 33
T2 sagital TSE	FOV de 220 mm TR: 5000 TE: 85 Nex: 1 Matrix: $0,7 \times 0,7 \times 3,5$ mm Espesor de corte: 3,5 mm Cortes: 25
DWI	Valores de b de 50, 500 y 1400 con sus respectivos mapas de ADC FOV de 280 mm TR: 5000 TE: 83 Nex: 1 Matrix: $0,9 \times 0,9 \times 3,5$ mm Espesor de corte: 3,5 mm Cortes: 28
T1 secuencias dinámicas	FOV de 260 mm TR: 4,46 TE: 1,64 Nex: 1 Matrix: $1,4 \times 1,4 \times 3,5$ mm Espesor de corte: 3,5 mm Mediciones: 23

ADC: Apparent Diffusion Coefficient; DWI: Diffusion-Weighted Imaging; FOV: Field Of View; TSE: Turbo Spin Echo.

el sistema de almacenamiento y comunicación (PACS, *Picture Archiving and Communication System*) de imágenes médicas de una institución de imágenes diagnósticas de la ciudad de Medellín (Colombia), previa aprobación por el comité de ética en investigación de la institución. Se incluyeron los estudios de RM de próstata de pacientes con sospecha de carcinoma prostático por alteración en el tacto rectal o elevación del PSA (estudio ordenado por su médico tratante). Se excluyeron los estudios incompletos o de mala calidad (definidos como imágenes que presentaran artefactos de movimiento o metálicos por la presencia de

**Figura 1.** Flujograma de la selección de pacientes.

implantes protésicos) y los estudios de pacientes que no presentaran esta sospecha diagnóstica (Fig. 1).

La selección de los estudios estuvo a cargo del investigador principal y de uno de los coinvestigadores, quienes no participaron en el posterior análisis de las imágenes. La recolección de la información se realizó en un formato (cuadro de Microsoft Excel) en donde los investigadores definieron si las imágenes eran evaluables y cumplían las características para entrar al estudio. Estos mismos investigadores revisaron los reportes de las RM originales realizados en el trabajo rutinario y se extrajo la puntuación PI-RADS allí consignada. Este puntaje fue utilizado para el análisis de concordancia intraobservador, al compararse la lectura previa ya realizada por cada uno de los observadores con la nueva evaluación de las imágenes por el mismo investigador que participó en el análisis inicial.

A los investigadores encargados de la puntuación de las imágenes con el PI-RADS v. 2.1 de los pacientes elegidos les fue entregado un listado anonimizado (seleccionado por los investigadores que no realizaron el análisis de las imágenes), con las imágenes de RM de los pacientes a evaluar, acompañado de una herramienta de recolección. Esta herramienta contenía en la primera columna el serial del paciente, en las columnas subsecuentes las principales características del PI-RADS v. 2.1 (T2, DWI, ADC, DCE) y en la última columna un espacio para el puntaje final de calificación.

Los tres investigadores involucrados en la evaluación de las imágenes son radiólogos subespecializados en imagen corporal con diferentes años de experiencia en la evaluación de imágenes de próstata, y específicamente usando el PI-RADS para su clasificación (12 años el observador 1, 11 años el observador 2 y 9 años el observador 3). Estos revisaron los estudios que

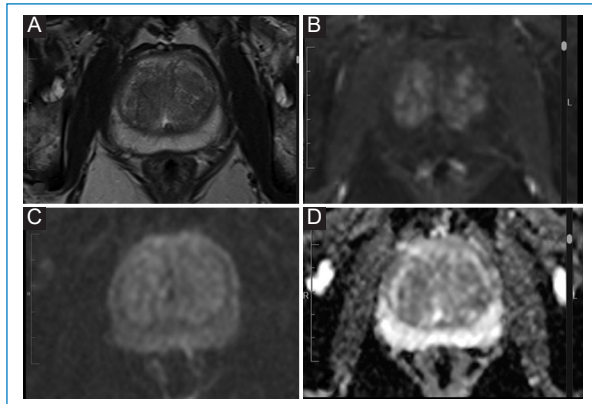


Figura 2. Adquisiciones axiales en T2 (A), T1 poscontraste (B), DWI (C) y ADC (D). El tejido glandular periférico demuestra intensidad de señal homogénea en T2, facilita la difusión y no demuestra zonas de realce patológico. Categoría PI-RADS 1.

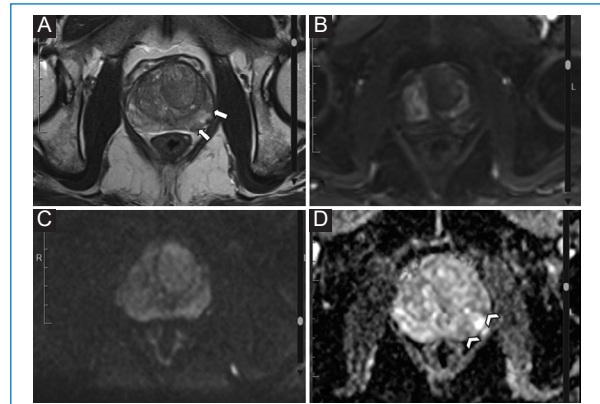


Figura 4. Adquisiciones axiales en T2 (A), T1 poscontraste (B), DWI (C) y ADC (D). En el tejido glandular periférico de la base del lóbulo izquierdo se identifican dos zonas nodulares de baja señal en T2 (flechas), que demuestran valores bajos de ADC (puntas de flecha), pero sin hiperintensidad en DWI; no muestran realce focal.

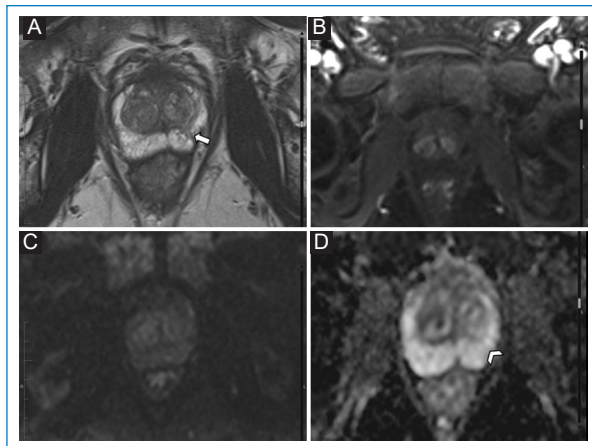


Figura 3. Adquisiciones axiales en T2 (A), T1 poscontraste (B), DWI (C) y ADC (D). Se identifica una zona de baja señal en T2 (flecha) ubicada en el tejido glandular periférico del lóbulo izquierdo, cuneiforme, sin restricción de la difusión y con leve reducción del coeficiente de atenuación aparente (punta de flecha); no muestra realce focal. Categoría PI-RADS 2.

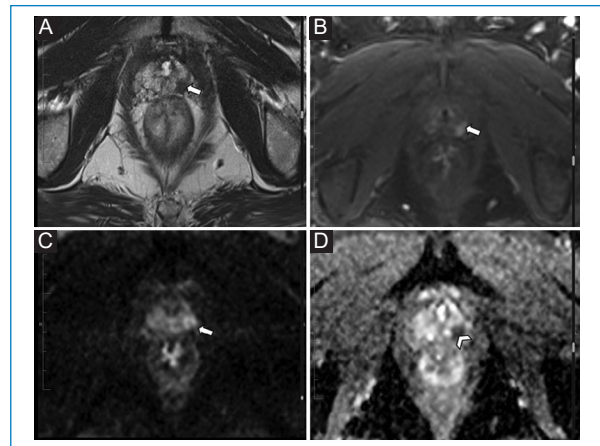


Figura 5. Adquisiciones axiales en T2 (A), T1 poscontraste (B), DWI (C) y ADC (D). En el tejido glandular periférico del lóbulo izquierdo se observa un nódulo sólido (flecha) de muy baja señal en T2, con realce focal poscontraste (flecha) y restricción en DWI (flecha) con su hipointensidad correspondiente en el mapa de ADC (punta de flecha). La lesión mide 8 mm y no presenta extensión extraprostática. Categoría PI-RADS 4.

cumplieron los criterios de inclusión (paciente con sospecha de carcinoma de próstata y RM con adecuados parámetros para su evaluación) mediante un visor de imágenes médicas (Hiruko, IMEXHS, Bogotá, Colombia). Calificaron las imágenes con los criterios principales del PI-RADS v. 2.1, consignaron los resultados en la herramienta de recolección descrita y emitieron una clasificación final para cada caso. En las Figs. 2 a 6 se muestran ejemplos de las imágenes obtenidas y evaluadas según la clasificación PI-RADS v. 2.1.

Los valores del índice kappa fueron clasificados de acuerdo con la tabla de Landis y Koch (1977) de este modo: 0,00 sin acuerdo; 0,01-0,20 leve; 0,21-0,40 aceptable; 0,41-0,60 moderada; 0,61-0,80 sustancial; y 0,81-1,00 casi perfecta.

En el presente estudio no se incluyó la correlación del puntaje PI-RADS con la histología, ya que el centro de diagnóstico donde se evaluaron las imágenes es de carácter ambulatorio. La falta de confirmación histológica no permite determinar con exactitud el valor de la interpretación.

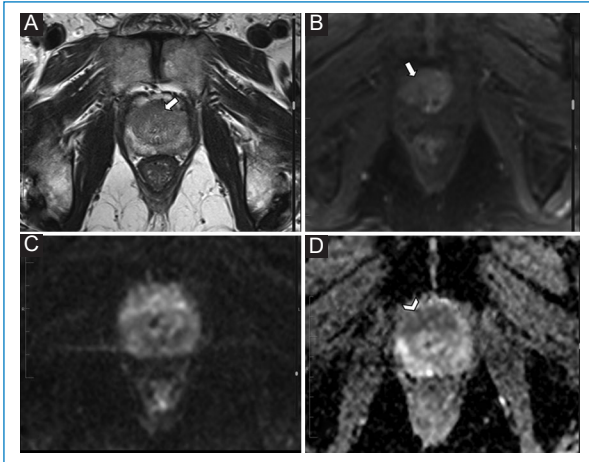


Figura 6. Adquisiciones axiales en T2 (A), T1 poscontraste (B), DWI (C) y ADC (D). Lesión sólida en el tejido glandular transicional anterior del ápex prostático, con compromiso bilateral, baja señal en T2 (flecha) y de forma lenticular, no encapsulado, que demuestra restricción de la difusión (con valores ADC muy bajos, punta de flecha) y con realce anormal poscontraste (flecha). Mide 25 mm en el eje transversal. Categoría PI-RADS 5.

Procesamiento de la información

Para describir las características de los pacientes se utilizaron medidas de resumen y de tendencia central para la edad y el PSA, previa verificación de su distribución normal con la prueba de Kolmogórov-Smirnov.

Para determinar la concordancia intra- e interobservador se calculó el estadístico kappa ponderado con pesos cuadráticos, con su respectivo IC95%, el estadístico Z y el valor p para la significancia estadística.

Las categorías del PI-RADS se agruparon en tres, uniando los calificados como 1 y 2, al igual que los calificados como 4 y 5, y se calculó igualmente la concordancia inter- e intraobservador teniendo en cuenta los años de experiencia de los radiólogos. Todos los análisis fueron realizados con TIMi Suite®.

Resultados

Características de los pacientes

Este estudio contó con la aprobación del comité de ética institucional. En total fueron incluidos en el estudio 129 pacientes mayores de 18 años a quienes se realizó RM de próstata, que fue leída por tres observadores en dos momentos del tiempo. La edad promedio fue de 64 (± 8) años, con límites de 42 y

86 años. Tenían registro del PSA 76 (58,9%) pacientes, con un valor mínimo de 1,37 ng/ml y máximo de 86 ng/ml, con un promedio de 7,49 ng/ml ($\pm 9,11$) (Fig. 7).

Reproducibilidad del PI-RADS v. 2.1: concordancia interobservador

La concordancia interobservador fue significativa para las categorías 2, 4 y 5 del PI-RADS, en las cuales se encontraron valores kappa de 0,555, 0,556 y 0,686, respectivamente, siendo la categoría 5 la de mayor acuerdo. En la categoría PI-RADS 1 se obtuvo un valor kappa negativo ($-0,008$), lo que significa discordancia, pero con un valor p estadísticamente no significativo, al igual que en la categoría PI-RADS 3.

La concordancia global fue moderada y significativamente distinta de la esperada por el azar, con un valor $p < 0,01$ (Tabla 2).

La concordancia interobservador de la categoría PI-RADS 3 fue de 0,077, con un intervalo de confianza entre $-0,075$ y $0,229$, y con un valor p no significativo (0,129). Esto puede estar explicado por la poca prevalencia de la categoría en la investigación (Tabla 2).

Agrupar el PI-RADS en los valores 1 y 2 (puntaje benigno), 3 (indeterminado) y 4 y 5 (puntaje maligno) da la posibilidad de un valor estadísticamente más significativo, sin tener esta agrupación una diferencia importante en la intervención clínica (Tabla 3).

Cuando se realizó el análisis según los años de experiencia de los radiólogos, la concordancia interobservador fue significativa en todos los casos, con valores de $p < 0,01$. Además, fue sustancial ($> 0,6$) entre todos los observadores cuando el PI-RADS se analizó con sus cinco categorías (Tabla 4).

El mayor acuerdo entre evaluadores se dio entre el observador 1 (con 12 años de experiencia) y el observador 3 (con 9 años de experiencia), con un valor kappa de 0,777, concordancia significativamente diferente de la esperada por el azar (Tabla 4).

Reproducibilidad test-retest del PI-RADS v. 2.1: concordancia intraobservador

La mayor reproducibilidad la tuvo el observador 3 (con 9 años de experiencia), al obtener el valor kappa intraobservador más alto, de 0,856 (IC95%: 0,777-0,935), y valores p significativos (Tabla 5).

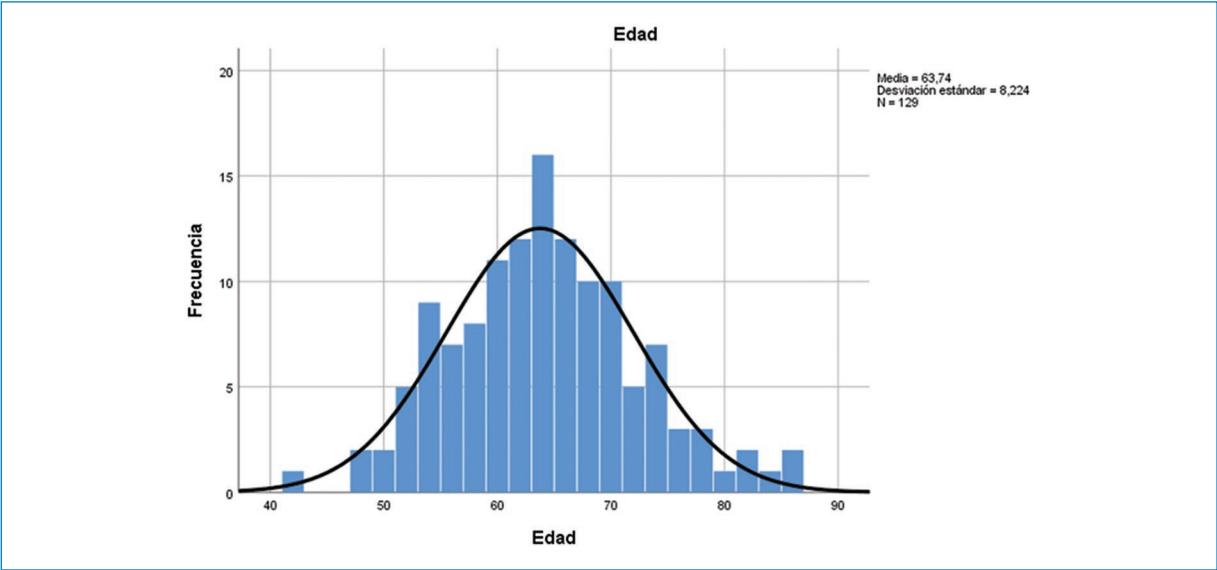


Figura 7. Histograma de frecuencias de la edad de los pacientes a quienes se realizó RM multiparamétrica de la próstata para detección de cáncer.

Tabla 2. Concordancia interobservador de las cinco categorías del PIRADS v. 2.1

PI-RADS	Kappa (IC95%)	p
1	-0,008 (-0,017-0,001)	0,878
2	0,555 (0,430-0,679)	0,000
3	0,077 (-0,075-0,229)	0,129
4	0,556 (0,374-0,738)	0,000
5	0,686 (0,524-0,848)	0,000
Kappa Global	0,556 (0,416-0,645)	0,000

IC95%: intervalo de confianza del 95%.

Tabla 3. Concordancia interobservador de las categorías del PI-RADS agrupadas en benigna (1 y 2), indeterminada (3) y maligna (4 y 5)

PI-RADS	Kappa (IC95%)	p
1 y 2	0,573 (0,448-0,698)	0,000
3	0,077 (-0,075-0,229)	0,129
4 y 5	0,679 (0,559-0,800)	0,000
Kappa Global	0,570 (0,452-0,689)	0,000

IC95%: intervalo de confianza del 95%.

El observador 2 (con 12 años de experiencia) obtuvo un valor kappa > 0,6, que se considera sustancial, y

Tabla 4. Concordancia interobservador del PIRADS v. 2.1 según la experiencia de los observadores

PI-RADS	Observadores*	Kappa (IC95%)	p
Cinco categorías	Observador 1 vs. 2	0,668 (0,513-0,822)	0,000
	Observador 2 vs. 3	0,663 (0,517-0,810)	0,000
	Observador 1 vs. 3	0,777 (0,677-0,878)	0,000

IC95%: intervalo de confianza del 95%.
*Observador 1: 12 años de experiencia; observador 2: 11 años de experiencia; observador 3: 9 años de experiencia.

Tabla 5. Concordancia intraobservador del PIRADS v. 2.1 según la experiencia de los observadores

PI-RADS	Observadores*	Kappa (IC95%)	p
Cinco categorías	Observador 1	0,781 (0,613-0,949)	0,000
	Observador 2	0,039 (-0,393-0,472)	0,875
	Observador 3	0,856 (0,777-0,935)	0,000

IC95%: intervalo de confianza del 95%.
*Observador 1: 12 años de experiencia; observador 2: 11 años de experiencia; observador 3: 9 años de experiencia.

su concordancia fue significativamente diferente de la esperada por el azar (Tabla 5).

Al analizar los resultados del observador 2 (11 años de experiencia) se observa que los valores kappa estuvieron por debajo de lo considerado aceptable; sin embargo, los

valores p fueron no significativos, que dan cuenta de una baja prevalencia de estudios interpretados en la base de datos original. Por lo tanto, este valor de concordancia intraobservador no puede ser tenido en cuenta (Tabla 5).

Discusión

En cuanto a la concordancia interobservador, en el presente estudio se obtuvo un kappa global de 0,556 (concordancia moderada) y un kappa para la categoría PI-RADS 5 de 0,686 (sustancial). Estos datos soportan la reproducibilidad del método, en especial cuando se trata de la detección de lesiones con mayor probabilidad de cáncer clínicamente significativo.

En las categorías PI-RADS 1 y 3 se encontraron concordancias interobservador sin acuerdo y leve, respectivamente, pero sin significancia estadística ($p > 0,05$). Lo anterior puede explicarse por el efecto negativo que tiene la baja prevalencia de un evento determinado sobre el estadístico kappa.

Cuando se realizó la subdivisión en tres categorías (benigno, indeterminado y maligno) se obtuvo concordancia moderada para las categorías con mayor probabilidad de ser benignas (1 y 2), y concordancia sustancial para aquellas con la mayor probabilidad de ser malignas (4 y 5). Lo anterior habla sobre la utilidad que tienen estos sistemas de puntuación de categorizar los hallazgos y dirigir la toma de decisiones clínicas, de tal manera que los hallazgos benignos retornen a un seguimiento clínico convencional y los malignos a la realización de tratamientos dirigidos. En la categoría 3 se encontró una p no significativa, probablemente por la poca proporción de pacientes, por lo que, en el presente estudio, no se hace énfasis en su resultado por su nula validez estadística.

En el estudio de Rosenkrantz et al.⁹ se encontraron valores kappa casi iguales a los encontrados en nuestro estudio, con 0,552 para las categorías PI-RADS > 4 , lo que indica la reproducibilidad de los criterios PI-RADS entre diferentes instituciones.

En el estudio de Labra et al.¹⁰ se obtuvieron niveles de concordancia global sustanciales (kappa 0,69) y además se reportó una mayor concordancia entre los observadores más experimentados. En el estudio de Smith et al.⁸ se encontró una concordancia intraobservador moderada (kappa 0,43 y 0,54) para observadores con poca experiencia, y sustancial (kappa 0,67) para un observador experto. En nuestro estudio, al comparar la concordancia según los niveles de experiencia, esta fue más alta entre los extremos. Al comparar el observador de mayor experiencia con el menos experimentado, el valor de kappa fue de 0,777, con $p < 0,05$, cerca del límite superior de sustancial.

Este resultado puede ser contraintuitivo, pero probablemente indica que la apropiación personal de los conceptos consignados en el PI-RADS puede ser más importante que la experiencia misma. En nuestro estudio, la concordancia intraobservador fue sustancial para el observador con mayor experiencia y casi perfecta para el observador con menor experiencia.

Otros estudios han investigado la concordancia intraobservador con otros sistemas de puntuación, como el de Vásquez et al.¹¹ con TI-RADS (*Thyroid Imaging Reporting and Data System*), en el que se reportó el mayor nivel de concordancia en el observador menos experimentado (un residente de Radiología). Los hallazgos de nuestro estudio y del de Vásquez et al.¹¹ van en contra de otros estudios que evalúan la concordancia, la cual usualmente aumenta con el nivel de experiencia¹².

En el estudio que se presenta se evalúa la variabilidad inter- e intraobservador con PI-RADS v. 2.1, a diferencia de otros estudios previos en los que se realizó la evaluación de las imágenes de RM de próstata con la versión 2.0⁷⁻¹⁰. Una de las razones para la actualización de PI-RADS fue mejorar la concordancia entre los observadores, ya que se identificaron varios criterios de evaluación que requerían ajustes⁶.

La evaluación de estas mejoras se ha realizado en diferentes publicaciones. Por ejemplo, en el artículo de Bhayana et al.¹³ se evaluó el acuerdo interobservador en la zona periférica y de transición entre las versiones 2 y 2.1 del PI-RADS, observándose que se logró una mejor concordancia en la zona periférica, pero no en la zona transicional.

En nuestro estudio, en comparación con los resultados de estudios que evaluaron la concordancia de la versión 2.0⁷⁻¹⁰, se encontraron unos valores kappa muy similares y no mayores que lo esperado para la versión 2.1; sin embargo, estos hallazgos podrían explicarse por la novedad de la actualización.

Nuestro estudio tiene varias limitaciones. La naturaleza retrospectiva no permite el control ni la estandarización estricta sobre los protocolos de imagen obtenidos; no obstante, durante el periodo de estudio todos los exámenes fueron realizados con el mismo resonador y por el mismo grupo de tecnólogos, lo cual puede proveer algún grado de estandarización.

El nivel de concordancia intraobservador para el observador 2 y la concordancia interobservador para las categorías PI-RADS 1 y 3 fueron excesivamente bajos, lo cual se explica por una limitante bien conocida del estadístico kappa, que se afecta por la prevalencia del hallazgo bajo consideración. Para hallazgos poco prevalentes, un valor kappa bajo no necesariamente refleja

una mala concordancia interobservador. Esta limitante puede ser mejorada discretamente utilizando el kappa ponderado, el cual asigna más peso a las categorías de los extremos, o de mayor importancia; por ejemplo, PI-RADS 2 vs. 5 (hallazgo con mayor probabilidad de benignidad vs. mayor probabilidad de malignidad)¹⁴.

La forma en que se estableció la concordancia intra-observador también explica la baja prevalencia de estudios para el observador 2. Como se realizó una búsqueda retrospectiva de estudios comprendidos en un periodo de tiempo, no fue posible controlar quién interpretó dichos estudios originalmente. Por motivos ajenos, el observador 2 tuvo la menor cantidad de estudios interpretados entre enero y diciembre de 2019, y ello resultó en una concordancia intraobservador no controlable.

Nuestro estudio no incluyó la correlación del puntaje PI-RADS con la histología, a diferencia de otros estudios de la literatura¹⁵. Esto principalmente se debe a que el centro de diagnóstico donde se realizó el estudio es de carácter ambulatorio, y la falta de confirmación histológica no permite determinar con exactitud el valor de la interpretación.

Conclusiones

Los resultados obtenidos en el presente estudio indican que el PI-RADS v. 2.1 es un sistema reproducible con una concordancia global buena (moderada) para las diferentes categorías y con distintos niveles de experiencia.

Financiamiento

Los autores declaran no haber recibido financiamiento para este estudio.

Conflicto de intereses

Los autores declaran no tener conflicto de intereses.

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 han obtenido la aprobación del Comité de Ética para el análisis y publicación de datos clínicos obtenidos de forma rutinaria. El consentimiento informado de los pacientes no fue requerido por tratarse de un estudio observacional retrospectivo.

Bibliografía

1. Sistema General de Seguridad Social en Salud. Guía de práctica clínica (GPC) para la detección temprana, seguimiento y rehabilitación del cáncer de próstata. 2013. Guía No. GPC-2013-21. Disponible en: https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/INEC/IETS/GPC_Comple_Prostata.pdf.
2. Grossman DC, Curry SJ, Owens DK, Bibbins-Domingo K, Caughey AB, Davidson KW, et al. Screening for prostate cancer: US Preventive Services Task Force Recommendation Statement. JAMA. 2018;319:1901-13.
3. Serefoglu EC, Altinova S, Ugras NS, Akincioglu E, Asil E, Balbay MD. How reliable is 12-core prostate biopsy procedure in the detection of prostate cancer? Can Urol Assoc J. 2013;7:E293-8.
4. Kasivisvanathan V, Rannikko AS, Borghi M, Panebianco V, Mynderse LA, Vaarala MH, et al. MRI-targeted or standard biopsy for prostate-cancer diagnosis. N Engl J Med. 2018;378:1767-77.
5. Ahmed HU, El-Shater Bosaily A, Brown LC, Gabe R, Kaplan R, Parmar MK, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet. 2017;389:815-22.
6. PI-RADS®. Prostate Imaging – Reporting and Data System. 2019. Version 2.1. Disponible en: <https://www.acr.org/-/media/ACR/Files/RADS/PI-RADS/PI-RADS-V2-1.pdf>.
7. Woo S, Suh CH, Kim SY, Cho JY, Kim SH. Diagnostic performance of Prostate Imaging Reporting and Data System Version 2 for detection of prostate cancer: a systematic review and diagnostic meta-analysis. Eur Urol. 2017;72:177-88.
8. Smith CP, Harmon SA, Barrett T, Bittencourt LK, Law YM, Shebel H, et al. Intra- and interreader reproducibility of PI-RADS v2: a multireader study. J Magn Reson Imaging. 2019;49:1694-703.
9. Rosenkrantz AB, Ginocchio LA, Cornfield D, Froemming AT, Gupta RT, Turkbey B, et al. Interobserver reproducibility of the PI-RADS version 2 lexicon: a multicenter study of six experienced prostate radiologists. Radiology. 2016;280:793-804.
10. Labra WA, Fuente-Alba CS, Schiappacasse FG, Barahona ZD, Skoknic BV. Concordancia interobservador para la validación del PI-RADS V2. Rev Arg Radiol. 2019;83:49-55.
11. Vázquez CMG, Durán JAM, Zapata SI, Londoño JFG, Gómez VG. Concordancia del TIRADS-ACR. Radiología. 2021;63:469-75.
12. Koh J, Kim SY, Lee HS, Kim EK, Kwak JY, Moon HJ, et al. Diagnostic performances and interobserver agreement according to observer experience: a comparison study using three guidelines for management of thyroid nodules. Acta Radiol. 2018;59:917-23.
13. Bhayana R, O'Shea A, Anderson MA, Bradley WR, Gottumukkala RV, Mojtaba A, et al. PI-RADS versions 2 and 2.1: interobserver agreement and diagnostic performance in peripheral and transition zone lesions among six radiologists. Am J Roentgenol. 2021;217:141-51.
14. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005;37:360-3.
15. Girometti R, Giannarini G, Greco F, Isola M, Cereser L, Como G, 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;49:546-55.