



ORIGINAL

[Artículo traducido] Acuerdo inter- e intraobservador independiente de la clasificación de Schizas de estenosis lumbar degenerativa. Comparación entre tres niveles de entrenamiento quirúrgico

F. Holc^a, A. Albani-Forneris^a, G. Kido^a, S. Beltrame^a, M. Petracchi^a, M. Gruenberg^a, C. Sola^a y G. Camino-Willhuber^{a,b,*}

^a Orthopaedic and Traumatology Department, Institute of Orthopedics Carlos E. Ottolenghi, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina

^b Department of Orthopaedics, University of California at Irvine, Orange, CA, Estados Unidos

Recibido el 29 de junio de 2022; aceptado el 1 de octubre de 2022

Disponible en Internet el 28 de febrero de 2023

PALABRAS CLAVE

Acuerdo independiente;
Clasificación de Schizas;
Estenosis espinal lumbar

Resumen

Introducción y objetivos: La estenosis espinal lumbar es una condición frecuente relacionada con la edad que afecta la calidad de vida. Se han desarrollado múltiples clasificaciones para cuantificar la gravedad de la estenosis que afectan la comparación entre estudios y la comunicación homogénea entre cirujanos e investigadores. A pesar de que esta clasificación no ha mostrado una correlación directa con la clínica, la clasificación de Schizas parece ser un método simple para evaluar la estenosis. Nuestro objetivo fue evaluar el acuerdo independiente inter- e intraobservador de la clasificación de Schizas en la severidad de la estenosis. Además, comparamos la concordancia entre tres niveles de formación en cirugía de columna. **Materiales y métodos:** Se llevó a cabo un acuerdo independiente inter- e intraobservador entre los residentes ortopédicos principiantes, avanzados y los cirujanos de columna; 90 niveles lumbares de 30 pacientes fueron evaluados por 16 observadores. Se utilizó concordancia a través del kappa ponderado.

Resultados: La concordancia global interobservador e intraobservador fue de 0,57 (IC 95% = 0,52-0,63) y 0,69 (0,55-0,79), respectivamente. La concordancia interobservador según el nivel de formación arroja valores de 0,53 (0,46-0,60) para los residentes menores, 0,61 (0,54-0,67) para los residentes mayores y 0,67 (0,59-0,74) para los asistentes. La concordancia intraobservador fue de 0,54 (0,48-0,60) para principiantes, 0,60 (0,55-0,66) para avanzados y 0,66 (0,60-0,72) para cirujanos de columna.

Véase contenido relacionado en DOI:

<https://doi.org/10.1016/j.recot.2022.10.003>

* Autor para correspondencia.

Correo electrónico: gaston.camino@hospitalitaliano.org.ar (G. Camino-Willhuber).

<https://doi.org/10.1016/j.recot.2023.02.011>

1888-4415/© 2022 SECOT. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

KEYWORDS

Independent agreement;
Schizas's classification;
Lumbar spinal stenosis

Conclusión: La clasificación de Schizas mostró concordancia interobservador moderada y concordancia intraobservador sustancial. Entre los cirujanos de columna, se observó un acuerdo sustancial inter- e intraobservador. La clasificación permitió una comunicación aceptable entre los cirujanos de columna entrenados.

© 2022 SECOT. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Independent inter and intra-observer agreement of the Schizas's classification of degenerative lumbar stenosis: Comparison among three levels of surgical training

Abstract

Introduction and objectives: Lumbar spinal stenosis is a common age-related condition that affects the quality of life. Multiple classifications have been developed to quantify the severity of stenosis affecting comparison between studies and homogenous communication among surgeons and researchers. Even though this classification has not shown a direct clinical correlation, Schizas's classification appears to be a simple method to assess stenosis. Our objective was to evaluate the inter and intraobserver independent agreement of the Schizas's classification to assess stenosis severity. Additionally, we aimed to compare agreement among three levels of training in spine surgery.

Materials and methods: An independent inter and intra observer agreement was conducted among junior, senior orthopedic residents and attending spine surgeons. Ninety lumbar levels from 30 patients were evaluated by 16 observers. Weighted kappa agreement was used.

Results: Overall interobserver and intraobserver agreement was of 0.57 (95% CI = 0.52–0.63) and 0.69 (0.55–0.79), respectively. Interobserver agreement according to level of training yielded values of 0.53 (0.46–0.60) for junior residents, 0.61 (0.54–0.67) for senior residents and 0.67 (0.59–0.74) for attendings. Intraobserver agreement was of 0.54 (0.48–0.60) for junior, 0.60 (0.55–0.66) for senior and 0.66 (0.60–0.72) for attendings.

Conclusion: The Schizas's classification showed moderate interobserver and substantial intra-observer agreement. Among attending surgeons, substantial inter and intraobserver agreement was observed. The classification allowed acceptable communication among trained spine surgeons.

© 2022 SECOT. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

La estenosis espinal lumbar (LSS) es una de las situaciones degenerativas más comunes de la columna, siendo la primera causa de cirugía espinal en personas mayores de 65 años de edad^{1–3}.

La presentación clínica implica claudicación neurogénica⁴, la cual empeora con la extensión lumbar y mejora con la flexión, afectando a las actividades ambulatorias⁵. Junto con los síntomas clínicos, la imagen de resonancia magnética (IRM) se ha convertido en esencial para la confirmación diagnóstica y la toma de decisiones quirúrgica⁶.

Se ha sugerido el área de superficie transversal para el diagnóstico y la severidad, aunque este sistema adolece de correlación clínica⁷. En 2010, Schizas et al.⁸ propusieron una clasificación de 7 categorías basada en el aspecto morfológico del saco dural en las imágenes axiales de RM de la columna lumbar ponderadas en T2, analizando la disposición del líquido cefalorraquídeo (LCR) y las raíces nerviosas, que tiene la ventaja de no requerir herramientas de medición específicas, y de ser fácilmente aplicable a la práctica

diaria. Los autores reportaron una confiabilidad interobservador sustancial en su estudio original.

Aun cuando esta clasificación ha sido utilizada en otros estudios, solo se han realizado dos estudios independientes de consenso^{9,10}. Otra evaluación independiente de la fiabilidad de este sistema es importante para validar adicionalmente su uso. Por tanto, el objetivo de este estudio es realizar un análisis de concordancia inter- e intraobservador independiente. Además, el objetivo de los autores fue comparar el acuerdo entre los observadores de diferentes niveles de formación en cuanto a enfermedad espinal-cirugía espinal.

Métodos

Se obtuvo aprobación del comité de revisión institucional para la realización del estudio, obviándose el consentimiento informado debido a la naturaleza retrospectiva del mismo, sin riesgo para los participantes. Seleccionamos retrospectivamente y al azar imágenes preoperatorias de RM lumbar de 30 pacientes, que fueron examinados en nuestro departamento clínico ambulatorio entre 2019 y 2020. La IRM

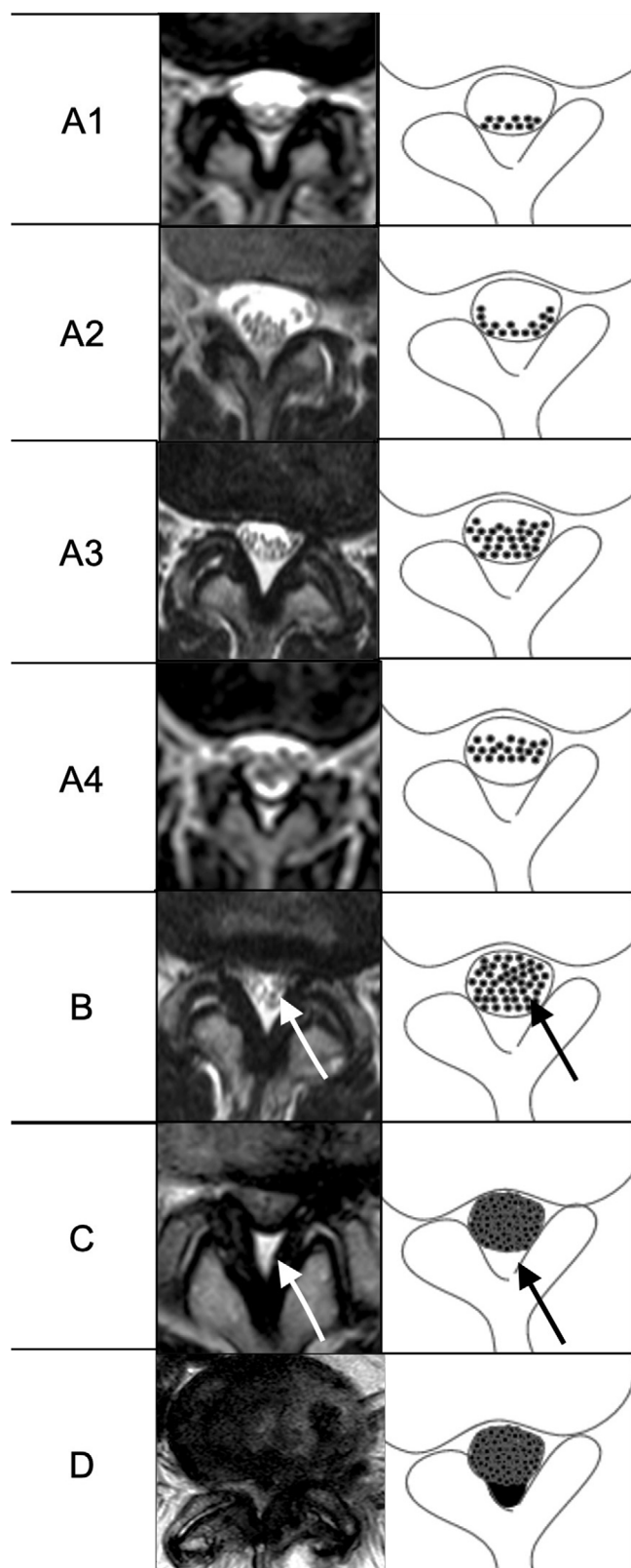


Figura 1 Representación de IRM y esquemática de la clasificación de Schizas para la estenosis lumbar. Las flechas de la imagen B muestran las radículas individualizadas que establecen la diferencia en cuanto a estenosis de grado B o moderada y estenosis de grado C o severa. Las flechas de la imagen C muestran la presencia de grasa epidural posterior que establece la

fue solicitada por diversos motivos, tales como dolor lumbar crónico, dolor en la pierna o claudicación neurogénica. Se excluyeron IRM del análisis en los pacientes con historia o sospecha de traumatismo, infección o tumor, cirugía lumbar previa y falta de imagen axial ponderada en T2 a cualquiera de los niveles de disco L2-L3, L3-L4, o L4-L5. Un autor con alto nivel de experiencia (M.P.), que posteriormente no participó en la fase de clasificación del estudio, seleccionó los casos. Todas las imágenes de RM se almacenaron digitalmente (DICOM) en la base de datos.

La clasificación de Schizas es un sistema de 7 grados basado en el ratio LCR/radicula en las imágenes axiales ponderadas en T2: el grado A (estenosis nula o menor), donde el LCR se observa en el interior del saco dural y su distribución es homogénea. A su vez, este grado se divide en 4 subgrupos, de acuerdo con la disposición de las raíces nerviosas, donde en A1 las raíces permanecen en la región dorsal y ocupan menos de la mitad del área del saco; A2 donde las raíces permanecen en la región dorsal y se distribuyen en herradura; A3 donde las raíces permanecen en la región dorsal y ocupan más de la mitad del área del saco dural; A4 donde las raíces se encuentran en la región central y ocupan la mayoría del área del saco dural. El grado B (estenosis moderada) abarca aquellas estenosis en las que las raíces ocupan la totalidad del área del saco dural, pero aún pueden individualizarse y existe una cierta cantidad de LCR entre ellas, aportando un aspecto granular. En el grado C (estenosis severa) las raíces no pueden reconocerse de manera individual, lo cual origina una señal homogénea gris dentro del saco y la grasa epidural posterior sigue presente; y el grado D (estenosis extrema) es equivalente al grado C, con la diferencia de que no puede visualizarse posteriormente grasa epidural (fig. 1).

El análisis de las imágenes fue realizado por 16 ortopedistas con tres niveles diferentes de experiencia. Tres eran residentes de posgrado de primer año (MIR-1) y 4 de segundo año (MIR-2) que formaron el primer grupo; 3 residentes de tercer año (MIR-3) y 3 de cuarto año (MIR-4) que formaron el segundo grupo; y 3 cirujanos de columna con al menos 5 años de experiencia en cirugía de columna, que formaron el tercer grupo.

Los evaluadores recibieron formación sobre este sistema de clasificación mediante una sesión online, a fin de debatirlo y aclarar dudas antes de realizar las evaluaciones, para estandarizar el proceso de evaluación. Se les suministró además el artículo original de Schizas et al.⁸ para resolver cualquier duda en el momento de la evaluación. Todos los participantes evaluaron de manera independiente el grado de estenosis del canal lumbar a los niveles de L2-L3, L3-L4, y L4-L5. Cada observador clasificó un total de 90 niveles. En la segunda fase, tras un intervalo de tiempo de 4 semanas para evitar el sesgo de memoria, todos los observadores realizaron el mismo examen con una distribución aleatoria de los casos.

diferencia entre estenosis de grado C o severa y estenosis de grado D o extrema.

Tabla 1 Concordancia interobservador de acuerdo al nivel de experiencia

Nivel de experiencia	Kappa de la primera evaluación (IC 95%)	Kappa de la segunda evaluación (IC 95%)
MIR-1 y 2	0,53 (0,46-0,60)	0,53 (0,46-0,59)
MIR-3 y 4	0,61 (0,54-0,67)	0,60 (0,53-0,67)
Adjuntos	0,67 (0,59-0,74)	0,60 (0,52-0,68)

Análisis estadístico

El análisis estadístico fue realizado utilizando el software Stata 16 (StataCorp, College Station, TX).

Considerando que la clasificación desarrollada por Schizas et al. es una variable ordinaria basada en la gravedad, decidimos utilizar las estadísticas kappa ponderadas (*wk*) para acuerdos bidireccionales. Kappa ponderada permite medir el acuerdo con múltiples niveles de respuesta, cuando no todos los desacuerdos son igualmente importantes. Se estableció la ponderación linealmente. El acuerdo interobservador fue determinado comparando la lectura inicial de todos los evaluadores. El acuerdo intraobservador fue calculado comparando las lecturas del mismo observador entre dos evaluaciones de los mismos pacientes. Se separaron las dos evaluaciones durante un intervalo de 4 semanas, presentándose en una secuencia aleatoria para evitar el sesgo de memoria.

Los niveles de concordancia para *wk* fueron determinados según lo propuesto por Landis et al.¹¹, del modo siguiente: valores *k* de 0 a 0,20, acuerdo leve; de 0,21 a 0,4, acuerdo bueno; de 0,41 a 0,6, acuerdo moderado; de 0,61 a 0,8, acuerdo sustancial; y de 0,81 a 1, acuerdo casi perfecto. Todos los acuerdos se expresan con un intervalo de confianza (IC) del 95%.

Resultados

Los acuerdos interobservador e intraobservador globales fueron 0,57 (IC 95%=0,52-0,63) y *k*=0,69 (0,55-0,79), respectivamente, correspondiendo a un acuerdo moderado y sustancial.

Concordancia interobservador de acuerdo al nivel de formación

La concordancia interobservador se ilustra en la [tabla 1](#). Sobre la base del nivel de formación, el acuerdo para los MIR 1-2 fue menor (0,53) en comparación con los MIR-3 y 4 (0,6) y los adjuntos (0,67).

Concordancia intraobservador de acuerdo al nivel de formación

Tanto MIR-1 y 2 y MIR-3 y 4 mostraron una concordancia moderada (0,54) en comparación con el acuerdo sustancial observado en los adjuntos (0,66) ([tabla 2](#)).

Tabla 2 Concordancia intraobservador de acuerdo al nivel de experiencia

Nivel de experiencia	Kappa	(IC 95%)
MIR-1 y 2	0,54	(0,48-0,60)
MIR-3 y 4	0,60	(0,55-0,66)
Adjuntos	0,66	(0,60-0,72)

Tabla 3 Concordancia interobservador de acuerdo al nivel de severidad de la estenosis lumbar

Estenosis	Kappa	(IC 95%)
Grado A1	0,59	(0,593-0,594)
Grado A2	0,26	(0,267-0,268)
Grado A3	0,41	(0,410-0,412)
Grado A4	0,20	(0,266-0,207)
Grado B	0,27	(0,279-0,28)
Grado C	0,39	(0,394-0,395)
Grado D	0,53	(0,533-0,534)

Concordancia interobservador de acuerdo a la severidad de la estenosis lumbar

Se observó un acuerdo bueno para los grados A2, A4 y B, mientras que los grados A1, A3, C y D mostraron un acuerdo moderado ([tabla 3](#)).

Concordancia intraobservador para cada grupo de acuerdo a la severidad de la estenosis lumbar

A2, A3 y D mostraron la misma concordancia respectivamente en todos los grupos, siendo buena para A2, y moderada para A3 y D. A1 y B tuvieron un mejor acuerdo en el grupo de adjuntos ([tabla 4](#)).

Discusión

La LSS es uno de los diagnósticos más frecuentes y una de las indicaciones más comunes en la cirugía de columna. Los análisis de concordancia independientes de cualquier clasificación son útiles para aportar validaciones externas, que son esenciales para facilitar la comunicación entre los facultativos, estandarizar la terminología de la investigación, y ayudar a guiar la toma de decisiones sobre los pacientes^{12,13}. En este estudio, validamos los acuerdos inter- e intraobservador de un sistema de estadificación de IRM para la LSS central entre observadores de diferentes niveles de formación. A este respecto, encontramos acuerdos interobservadores globales moderados y acuerdos intraobservadores sustanciales (0,57 [IC 95%=0,52-0,63] y 0,69 [0,55-0,79]), respectivamente. Nuestros resultados mostraron un mejor acuerdo en comparación con los de los autores originales de la clasificación⁸. Los valores medios de kappa inter- e intraobservadores de Schizas et al. fueron moderados (0,44) y sustanciales (0,65), respectivamente.

Ko et al.¹², en su trabajo de correlación, incluyeron 5 observadores, de los cuales 2 fueron individuos con 3 meses de formación, un residente de imagen diagnóstica con un

Tabla 4 Concordancia intraobservador para cada grupo de acuerdo al nivel de severidad de la estenosis lumbar

Estenosis	PGY-1 y 2		PGY-3 y 4		Adjuntos	
	Kappa	(IC 95%)	Kappa	(IC 95%)	Kappa	(IC 95%)
Grado A1	0,59	(0,518-0,608)	0,59	(0,534-0,641)	0,66	(0,544-0,783)
Grado A2	0,25	(0,209-0,299)	0,25	(0,198-0,304)	0,33	(0,209-0,448)
Grado A3	0,42	(0,374-0,465)	0,48	(0,423-0,530)	0,41	(0,288-0,526)
Grado A4	0,19	(0,144-0,234)	0,22	(0,172-0,279)	0,31	(0,190-0,429)
Grado B	0,31	(0,264-0,354)	0,27	(0,220-0,327)	0,45	(0,334-0,573)
Grado C	0,32	(0,279-0,369)	0,45	(0,393-0,500)	0,48	(0,363-0,602)
Grado D	0,51	(0,468-0,558)	0,51	(0,456-0,562)	0,59	(0,466-0,705)

mes de experiencia, y 2 fueron cirujanos con más de 10 años de experiencia. Ellos mostraron un índice de correlación intraclass (ICI) de 0,82-0,98, equivalente a un acuerdo casi perfecto. En cuanto a la correlación intraobservador, ellos reportaron un ICI de 0,65-0,99, lo cual se interpreta como un acuerdo de moderado a excelente. Sin embargo, estos resultados no son comparables con los obtenidos en nuestro estudio, dado que se utilizó otro método de correlación¹².

En otro estudio de correlación realizado por Weber et al.⁹, que incluyó 2 neurocirujanos y 2 especialistas en imagen diagnóstica, todos ellos con amplia experiencia en su campo, se obtuvo un nivel de acuerdo interobservador sustancial $k=0,76$ (IC 95% = 0,69-0,83) y un nivel de acuerdo intraobservador casi perfecto $k=0,76$ a 0,96. A diferencia de nuestro estudio, todos los observadores tenían amplia experiencia en el análisis de las imágenes de RM y, a su vez, utilizaron la clasificación abreviada de Schizas, dado que no se consideraron las subclasificaciones de grado A, lo cual pudo afectar por tanto al valor de estimación de la correlación, reduciendo una diferencia potencial entre las observaciones.

Lønne et al.¹⁰ reportaron un acuerdo interobservador sustancial $k=0,65$ (IC 0,56-0,74) entre 2 radiólogos experimentados que utilizaron este sistema, y un acuerdo intraobservador alto $k=0,78$ (IC 0,65-0,92) y $k=0,81$ (IC 0,68-0,94), respectivamente. Al igual que en el estudio realizado por Weber et al., los observadores eran especialistas experimentados y utilizaron la clasificación de Schizas abreviada, lo cual podría haber reflejado una contribución baja a la validez externa más allá de los especialistas en imagen diagnóstica. Además, todos los pacientes de este estudio fueron candidatos a la cirugía, lo cual pudo suponer una selección sesgada.

Al estratificar el análisis de acuerdo al nivel de formación, encontramos una concordancia inter- e intra-observador moderada entre los residentes MIR-1 y PGY-2 (0,53 y 0,54, respectivamente), mientras que MIR-3 y MIR-4 reflejaron un acuerdo interobservador sustancial (0,61) y moderado (0,6). Los resultados mayores se observaron en los residentes con mayor nivel de experiencia (interobservador = 0,67, intraobservador = 0,66). Dichas diferencias podrían explicarse mediante el hecho de que los residentes MIR-1 y MIR-2 están inherentemente expuestos a menores casos de IRM de columna, en comparación con los MIR-3 y MIR-4, ya que los residentes más jóvenes no evaluaban de manera rutinaria a los pacientes en la unidad de columna en dicho momento de la formación, y solo se expusieron a casos

de columna de manera eventual durante las actividades de las guardias¹³⁻¹⁵.

De manera interesante, al comparar la concordancia entre los residentes MIR-3 y MIR-4 y los médicos adjuntos, se observó una diferencia pequeña, en cuanto a que la capacidad de registrar y utilizar esta clasificación es esencial y similarmente adquirida y procesada en ambos grupos. Este hallazgo respalda la idea de la utilidad de esta clasificación a efectos de comunicación e investigación. Cabe destacar que los participantes en este estudio no estaban previamente familiarizados con este sistema de clasificación, lo cual podría haber influido en nuestros resultados.

Diversos estudios han analizado la fiabilidad de los sistemas de clasificación entre observadores con diferentes niveles de experiencia. Algunos de ellos demostraron una correlación positiva entre el acuerdo y la experiencia clínica^{13,16,17}, mientras que otros no encontraron diferencias en cuanto a la fiabilidad de los observadores, sobre la base del nivel de experiencia^{15,18}.

Los análisis de fiabilidad^{9,12,15,16} utilizaron normalmente participantes de diversos niveles de experiencia y formación, en un intento de analizar la generalizabilidad del estudio, emulando una situación más real de la práctica clínica diaria. Sin embargo, a nuestro entender, este es el primer estudio que evalúa la concordancia de esta clasificación de acuerdo a la experiencia clínica y el nivel de formación. Aun cuando Ko et al.¹² involucraron a observadores de diferentes niveles de experiencia, no se realizó ningún análisis comparativo entre cada uno de los observadores.

Finalmente, al analizar el acuerdo sobre la base de la gravedad de la estenosis lumbar, según la propuesta realizada por Schizas et al., encontramos una concordancia leve en los grados A2, A4 y B, y moderada en los grados A1, A3, C y D. Creemos que este acuerdo heterogéneo podría explicarse por el hecho de que esta clasificación tiene 7 grados, considerando especialmente que el grado A tiene 4 subtipos, y las diferencias entre ellos parecen ser leves. Por tanto, al estudiar la concordancia para el grado de la severidad de la estenosis de acuerdo al nivel de experiencia, encontramos que los grados extremos eran también los que reflejaron mejor acuerdo, presentando el grupo experimentado una concordancia homogénea entre los diferentes grados.

Nuestro estudio tiene limitaciones y fortalezas. En primer lugar, analizamos las imágenes de RM sin ninguna correlación clínica. El que los autores no conocieran si la información clínica podría influir en la evaluación por parte de los evaluadores, incluyendo los hallazgos clínicos, en nuestro estudio,

podría reproducir un escenario clínico más real de la práctica diaria.

En segundo lugar, solo incluimos a los participantes con formación en ortopedia, sin participación de especialistas en radiología o neurocirugía. A pesar de que esto podría afectar a nuestros resultados, los autores seleccionaron a cirujanos y residentes de diferentes niveles de formación para entender mejor la influencia del nivel formativo a la hora de clasificar la estenosis lumbar.

Por último, las imágenes facilitadas para el análisis incluyeron el conjunto completo de IRM, en el que cada observador tenía que localizar y evaluar los niveles solicitados, simulando un contexto del «mundo real», y permitiendo una evaluación más auténtica de las imágenes.

Conclusión

Este estudio validó de manera independiente el acuerdo inter- e intraobservador del sistema de clasificación morfológica de Schizas para la LSS central entre los residentes de ortopedia con diferentes niveles de formación y cirujanos de columna. El acuerdo interobservador fue moderado para los residentes jóvenes, y sustancial para los residentes de mayor edad y los cirujanos de columna adjuntos, y el acuerdo intraobservador fue sustancial solo para los cirujanos adjuntos. Aun cuando esta clasificación no ha reflejado una correlación clínica directa, el sistema basado en la morfología de Schizas podría permitir un lenguaje de comunicación homogénea entre los facultativos, con niveles de formación avanzados. La utilidad última de esta clasificación debería evaluarse en estudios prospectivos de cohortes de mayor tamaño.

Nivel de evidencia

Nivel de evidencia IV.

Financiación

El(los) autor(es) no han recibido respaldo financiero alguno para la investigación, autoría y/o publicación de este artículo.

Conflicto de intereses

El(los) autor(es) han declarado la ausencia de conflictos de interés potenciales con respecto a la investigación, autoría y/o publicación de este artículo.

Bibliografía

- Genevay S, Atlas SJ, Katz JN. Variation in eligibility criteria from studies of radiculopathy due to a herniated disc and of neurogenic claudication due to lumbar spinal stenosis: a structured literature review. *Spine*. 2010;35:803–11.
- Kalichman L, Cole R, Kim DH, Li L, Suri P, Guermazi A, et al. Spinal stenosis prevalence and association with symptoms: the Framingham study. *Spine J*. 2009;9:545–50.
- Katz JN, Harris MB. Clinical practice. Lumbar spinal stenosis. *N Engl J Med*. 2008;358:818–25.
- Deer T, Sayed D, Michels J, Josephson Y, Li S, Calodney AK. A review of lumbar spinal stenosis with intermittent neurogenic claudication: disease and diagnosis. *Pain Med*. 2019;20 Suppl 2:S32–44.
- Katz JN, Dalgas M, Stucki G, Katz NP, Bayley J, Fossel AH, et al. Degenerative lumbar spinal stenosis. Diagnostic value of the history and physical examination. *Arthritis Rheum*. 1995;38:1236–41.
- Lurie J, Tomkins-Lane C. Management of lumbar spinal stenosis. *BMJ*. 2016;352:h6234.
- Schönström N, Lindahl S, Willén J, Hansson T. Dynamic changes in the dimensions of the lumbar spinal canal: an experimental study in vitro. *J Orthop Res*. 1989;7:115–21, <http://dx.doi.org/10.1002/jor.1100070116>.
- Schizas C, Theumann N, Burn A, Tansey R, Wardlaw D, Smith FW, et al. Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on magnetic resonance images. *Spine*. 2010;35:1919–24.
- Weber C, Rao V, Gulati S, Kvistad KA, Nygaard ØP, Lønne G. Inter- and intraobserver agreement of morphological grading for central lumbar spinal stenosis on magnetic resonance imaging. *Global Spine J*. 2015;5:406–10, <http://dx.doi.org/10.1055/s-0035-1551651>.
- Lønne G, Ødegard B, Johnsen LG, Solberg TK, Kvistad KA, Nygaard ØP. MRI evaluation of lumbar spinal stenosis: is a rapid visual assessment as good as area measurement? *Eur Spine J*. 2014;23:1320–4.
- Landis JR, Richard Landis J, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–74, <http://dx.doi.org/10.2307/2529310>.
- Ko Y-J, Lee E, Lee JW, Park CY, Cho J, Kang Y, et al. Clinical validity of two different grading systems for lumbar central canal stenosis: Schizas and Lee classification systems. *PLoS One*. 2020;15:e0233633.
- Yoon RS, Koerner JD, Patel NM, Sirkin MS, Reilly MC, Liporace FA. Impact of specialty and level of training on CT measurement of femoral version: an interobserver agreement analysis. *J Orthop Traumatol*. 2013;14:277–81, <http://dx.doi.org/10.1007/s10195-013-0263-x>.
- Jin EH, Chung SJ, Lim JH, Chung GE, Lee C, Yang JI, et al. Training effect on the inter-observer agreement in endoscopic diagnosis and grading of atrophic gastritis according to level of endoscopic experience. *J Korean Med Sci*. 2018;33:e117, <http://dx.doi.org/10.3346/jkms.2018.33.e117>.
- Niemeyer T, Wolf A, Kluba S, Halm HF, Dietz K, Kluba T. Interobserver and intraobserver agreement of Lenke and King classifications for idiopathic scoliosis and the influence of level of professional training. *Spine (Phila Pa 1976)*. 2006;31:2103–8, <http://dx.doi.org/10.1097/01.brs.0000231434.93884.c9>.
- Wing N, Van Zyl N, Wing M, Corrigan R, Loch A, Wall C. Reliability of three radiographic classification systems for knee osteoarthritis among observers of different experience levels. *Skeletal Radiol*. 2021;50:399–405, <http://dx.doi.org/10.1007/s00256-020-03551-4>.
- Tzavellas AN, Kenanidis E, Potoupnis M, Pellios S, Tsiridis E, Sayegh F. Interobserver and intraobserver reliability of Salter-Harris classification of physeal injuries. *Hippokratia*. 2016;20:222–6.
- Karamian BA, Schroeder GD, Levy HA, Canseco JA, Benneker LM, Kandziora F, et al., AO Spine Sacral Classification Group Members. The influence of surgeon experience and subspecialty on the reliability of the AO spine sacral classification system. *Spine (Phila Pa 1976)*. 2021;46:1705–13, <http://dx.doi.org/10.1097/BRS.0000000000004199>.