Tesis por compendio de publicaciones

Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen

Programa doctorado

Biociencias: Biología y Clínica del Cáncer y Medicina Traslacional

Instituto Universitario de Biología Molecular y Celular del Cáncer

Autor
Estanislao Arana Fernández de Moya

Director Luis Alberto Pérez Romasanta
Tutor Rogelio González Sarmiento

Agradecimientos

A todos los que me han enseñado y me permiten seguir aprendiendo
A mi familia

En realidad, uno nunca sabe adónde nos llevará la enseñanza; quién la escuchará y, en cierto casos, quién será el que enseñe.

Daniel Mendelsohn. Una Odisea. p.377. Seix Barral, 2019

Índice

Estimado Coordinador del Programa de Doctorado:

Estanislao Arana Fernández de Moya con DNI: 22559639C, y alumno del programa de doctorado "Biociencias: Biología y Clínica del Cáncer y Medicina Traslacional "

Solicita que se tenga en consideración la información aportada en este documento con el objetivo de poder presentar la tesis con título "*Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen*" mediante el formato de compendio de artículos/publicaciones. La información aportada se corresponde con lo establecido en el Procedimiento para la presentación de la tesis doctoral en la Universidad de Salamanca en el Formato de Compendio de Artículos/Publicaciones:

A continuación se detallan los documentos adjuntos en esta solicitud:

- Página Inicial especificando que la tesis corresponde a un compendio de trabajos previamente publicados detallando para cada uno de ellos: referencia de la revista, editorial, DOI y afiliaciones de cada uno de los miembros autores.
- Autorización del director o codirector para la presentación de la tesis mediante el formato de compendio de artículos/publicaciones.
- Tema objeto de estudio (Introducción). Hipótesis de trabajo, objetivos, principales conclusiones
- Copia completa de las publicaciones originales que conformarán la Tesis Doctoral (artículos, capítulos de libro, libro o libros aceptados o publicados).
- Para cada uno de los 3 artículos presentados, un resumen en castellano en el cual se especifican: los objetivos de la investigación, la metodología utilizada, los resultados alcanzados, y las conclusiones finales.

Compendio de trabajos publicados

La tesis "Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen " corresponde a un compendio de trabajos previamente publicados. A continuación se detalla el nombre y afiliación de cada uno de los autores, la referencia completa de la revista o editorial, DOI, PMID y cuartil de cada uno de ellos en JCR.

1. Agreement in the assessment of metastatic spine disease using scoring systems.
Radiother Oncol. 2015 Apr;115(1):135-40.

Arana E(1), Kovacs FM(2), Royuela A(3), Asenjo B(4), Pérez-Ramírez U(5), Zamora J(6); Spanish Back Pain Research Network Task Force for the improvement of inter-disciplinary management of spinal metastasis.

Author information:
(1)Department of Radiology, Valencian Oncology Institute Foundation, Valencia, Spain; Research Institute in Health Services Foundation, Valencia, Spain; Spanish Back Pain Research Network, Kovacs Foundation, Palma de Mallorca, Spain. Electronic address: aranae@uv.es.
(2)Spanish Back Pain Research Network, Kovacs Foundation, Palma de Mallorca, Spain; Scientific Department, Kovacs Foundation, Palma de Mallorca, Spain.
(3)Spanish Back Pain Research Network, Kovacs Foundation, Palma de Mallorca, Spain; CIBER Epidemiology and Public Health (CIBERESP), Spain; Clinical Biostatistics Unit, Hospital Ramón y Cajal, IRYCIS, Madrid, Spain.
(4)Spanish Back Pain Research Network, Kovacs Foundation, Palma de Mallorca, Spain; Department of Radiology, Hospital Regional Universitario Carlos Haya, Málaga, Spain.
(5)Spanish Back Pain Research Network, Kovacs Foundation, Palma de Mallorca, Spain; Center for Biomaterials and Tissue Engineering, Universitat Politècnica de València, Valencia, Spain.
(6)Spanish Back Pain Research Network, Kovacs Foundation, Palma de Mallorca, Spain; CIBER Epidemiology and Public Health (CIBERESP), Spain; Clinical Biostatistics Unit, Hospital Ramón y Cajal, IRYCIS, Madrid, Spain; Barts and the London School of Medicine & Dentistry. Queen Mary University of London, London, UK.

DOI: 10.1016/j.radonc.2015.03.016. PMID: 25869337
Factor de impacto JCR-2015 : 4,817. Q1 Oncology. Q1: Radiology, Nuclear Medicine & Medical imaging

2. Agreement in Metastatic Spinal Cord Compression.
J Natl Compr Canc Netw. 2016 Jan;14(1):70-6.

Arana E(1)(2)(3), Kovacs FM(3)(4), Royuela A(3)(5)(6), Asenjo B(3)(7), Pérez-Ramírez Ú(3)(8), Zamora J(3)(5)(6)(9); Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis.

Author information:
(1)Department of Radiology, Valencian Oncology Institute Foundation, Valencia, Spain

(2)Research Institute in Health Services Foundation, Valencia, Spain
(3)Spanish Back Pain Research Network, Kovacs Foundation, Palma de Mallorca, Spain
(4)Scientific Department, Kovacs Foundation, Palma de Mallorca, Spain
(5)CIBER Epidemiology and Public Health (CIBERESP), Madrid, Spain
(6)Clinical Biostatistics Unit, Hospital Ramón y Cajal, IRYCIS, Madrid, Spain
(7)Department of Radiology, Hospital Regional Universitario Carlos Haya, Málaga, Spain
(8)Center for Biomaterials and Tissue Engineering, Universitat Politècnica de València, Valencia, Spain
(9)Barts and the London School of Medicine & Dentistry, Queen Mary University of London, London, United Kingdom

DOI: 10.6004/jnccn.2016.0008 . PMID: 26733556
Factor de impacto JCR-2016 : 4,675. Q2 Oncology.

3. **Spine Instability Neoplastic Score: agreement across different medical and surgical specialties.**

Spine J. 2016 May;16(5):591-9. doi: 10.1016/j.spinee.2015.10.006. Epub 2015 Oct 22.

Arana E(1), Kovacs FM(2), Royuela A(3), Asenjo B(4), Pérez-Ramírez Ú(5), Zamora J(6); Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis.

Author information:
(1)Department of Radiology, Valencian Oncology Institute Foundation, C/ Beltrán Báguena, 19, 46009 Valencia, Spain; Research Institute in Health Services Foundation, C/ San Vicente,112, 3, 46007 Valencia, Spain; Spanish Back Pain Research Network, Kovacs Foundation, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain. Electronic address: aranae@uv.es.
(2)Spanish Back Pain Research Network, Kovacs Foundation, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain; Scientific Department, Kovacs Foundation, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain.
(3)Spanish Back Pain Research Network, Kovacs Foundation, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain; CIBER Epidemiology and Public Health (CIBERESP), Av. Monforte de Lemos, 3-5. Pabellón 11. Planta 0, 28029 Madrid, Spain; Clinical Biostatistics Unit, Hospital Ramón y Cajal, IRYCIS. Ctra. Colmenar Km. 9.1, 28034 Madrid, Spain.
(4)Spanish Back Pain Research Network, Kovacs Foundation, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain; Department of Radiology, Hospital Regional Universitario Carlos Haya, Avda Carlos Haya s/n, 29010 Málaga, Spain.
(5)Spanish Back Pain Research Network, Kovacs Foundation, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain; Center for Biomaterials and Tissue Engineering, Universitat Politècnica de València, CPI Building (8E), F access, 1st floor, Cami de Vera, s/n, 46022 Valencia, Spain.
(6)Spanish Back Pain Research Network, Kovacs Foundation, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain; CIBER Epidemiology and Public Health (CIBERESP), Av. Monforte de Lemos, 3-5. Pabellón 11. Planta 0, 28029 Madrid, Spain; Clinical Biostatistics Unit, Hospital Ramón y Cajal, IRYCIS. Ctra. Colmenar Km. 9.1, 28034 Madrid, Spain; Barts and the London School of Medicine & Dentistry, Queen Mary University of London, Mile End Road, London E1 4NS, UK.

DOI: 10.1016/j.spinee.2015.10.006. PMID: 26471708 FI 2015 2,96 Q2

Factor de impacto JCR-2016 : 2,96. Q1 Orthopedics. Q2: Clinical Neurology

4. Metastatic Versus Osteoporotic Vertebral Fractures on MRI: A Blinded, Multicenter, and Multispecialty Observer Agreement Evaluation.

J Natl Compr Canc Netw. 2020 Mar;18(3):267-273.

Arana E(1)(2), Kovacs FM(2)(3), Royuela A(2)(4), Asenjo B(2)(5), Nagib F(2)(5), Pérez-Aguilera S(2)(6), Dejoz M(2)(7), Cabrera-Zubizarreta A(2)(8),García-Hidalgo Y(2)(9), Estremera A(2)(10); Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis.

Author information:
(1)Department of Radiology, Fundación Instituto Valenciano de Oncología, Valencia.
(2)Spanish Back Pain Research Network, Kovacs Foundation, Palma de Mallorca.
(3)Unidad de la Espalda Kovacs, Hospital Universitario HLA-Moncloa, Madrid.
(4)Clinical Biostatistics Unit, Instituto de Investigación Sanitaria Puerta de Hierro-Segovia de Arana, Madrid.
(5)Department of Radiology, Hospital Universitario Regional de Málaga, Málaga.
(6)Department of Radiology, Hospital de Manacor, Mallorca.
(7)School of Biomedical Engineering, Universitat Politècnica de Valencia,Valencia.
(8)Department of Radiology, Hospital de Galdakao, Galdakao, Bizkaia.
(9)Department of Radiology, Hospital Universitario Puerta de Hierro, Madrid; and.
(10)Department of Radiology, Hospital Son Llàtzer, Palma de Mallorca, Spain.

DOI: 10.6004/jnccn.2019.7367.PMID: 32135511
Factor de impacto JCR-2018 : 7,579. Q1 Oncology.

Autorización del Director

Salamanca 23 de marzo de 2020

Dr Luis Alberto Pérez Romasanta con DNI 16007972H profesor asociado ciencias de la Salud del Departamento de Ciencias Biomédicas y del Diagnóstico de la Facultad de Medicina

Hago constar:

Como director, autorizo que la tesis doctoral de Estanislao Arana Fernández de Moya con DNI 22559639C matriculado en el programa de doctorado "Biociencias: Biología y Clínica del Cáncer y Medicina Traslacional de la Universidad de Salamanca "Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen" se presente como compilación de artículos pulblicados. Además confirmamos que es el autor principal de los mismos, siendo estos:

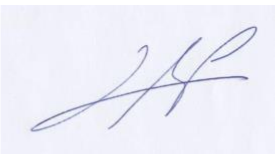
- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez U, Zamora J. Agreement in the assessment of metastatic spine disease using scoring systems. Radiother Oncol. 2015 Apr;115(1):135-40. Q1 Oncology

- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J. Agreement in Metastatic Spinal Cord Compression. J Natl Compr Canc Netw. 2016 Jan;14(1):70-6. Q2 Oncology. Q1 Oncology

- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J. Spine Instability Neoplastic Score: agreement across different medical and surgical specialties. Spine J. 2016 May;16(5):591-9. Q1 Orthopedics

- Arana E, Kovacs FM, Royuela A, Asenjo B, Nagib F, Pérez-Aguilera S, Dejoz M, Cabrera-Zubizarreta A,García-Hidalgo Y, Estremera A. Metastatic Versus Osteoporotic Vertebral Fractures on MRI: A Blinded, Multicenter, and Multispecialty Observer Agreement Evaluation. J Natl Compr Canc Netw. 2020 Mar;18(3):267-273. Q1 Oncology

Por todo ello firmo esta carta de autorización



Introducción

El esqueleto es el órgano donde más frecuentemente existen metástasis en pacientes con cáncer. Se estima que un 70% de pacientes oncológicos presentan metástasis vertebrales en su fase avanzada (1). Su diagnóstico indica un mal pronóstico, especialmente por la alta probabilidad de complicación esquelética relacionada (CER)(2). Las complicaciones más frecuentes son las fracturas vertebrales y la compresión medular. Estas complicaciones representa la mayor parte de los costes hospitalarios de estos pacientes en este estadio(1).

El tratamiento está marcado por el estadio de las lesiones vertebrales en conjunción con la patología general del paciente. Existen múltiples escalas de valoración de la enfermedad metastásica vertebral, tanto para el diagnóstico general, la inestabilidad vertebral y la compresión medular.

Los objetivos específicos del presente trabajo son aplicar el estudio del acuerdo intra- e interobservador a los siguientes aspectos clínicos:

- Diagnóstico y pronóstico de las metástasis vertebrales
- Valoración de la inestabilidad del raquis
- Valoración y estadio de compresión medular
- Diagnóstico diferencial entre fractura por osteoporosis y fractura por metástasis.

Respecto al diagnóstico y estadificación, la primera escala disponible fue Tomita y posteriormente surgió Tokushahi (3). Aunque ambas son precisas para determinar el pronóstico a 6 meses, esta última es mejor para el pronóstico a largo plazo según últimos estudios (4). No obstante han surgido recientes escalas que aparentemente sobrepasan las virtudes de otras, siendo la más reciente el normograma del grupo de investigación en oncología esquelética (SORG, en su acrónimo inglés), para la mortalidad a corto plazo (5). Existen constantes variaciones en el tipo y número de escalas escogidas (6). Se estima que una combinación de algoritmo manual e inteligencia artificial puede ser la mejor aproximación. Estas escalas serían más comprensibles cuando mostraran el resultado y la explicación del mismo, lo que haría el proceso más transparente. (6).

La compresión medular es la complicación más grave en la columna oncológica y se presenta hasta en un 55% de los pacientes con metástasis (7) y se considera generalmente uno de los signos para atención especializada oncológica a nivel mundial (8). Hasta hace poco no existían escalas de diagnóstico de la misma. Sólo recientemente se ha desarrollado y mostrado su fiabilidad por neurocirugía (9), "*Spinal cord compression scale*" (ESCC, en español). Con el avance de las técnicas de radioterapia es más necesario diferenciar los grupos diagnósticos y pronósticos (10). Así, aunque el mejor tratamiento se estima que es una cirugía agresiva seguida de radioterapia, no hay consenso en el esquema de radioterapia más eficaz posterior (10). El estudio del acuerdo interobservador entre todas las especialidades médicas involucradas en la es el objetivo de los artículos compendiados en esta tesis (11)

Existen también variedad de sistemas de escala y clasificación para la inestabilidad de los tumores en raquis. El problema es que no se ha probado su fiabilidad intra- ni interobservador antes de desarrollarlos. Entre ellos está la escala de inestabilidad de columna vertebral neoplásica (*SINS*, en su acrónimo inglés) (12), aunque existen otras (13). El avance de la cirugía de raquis tumoral con intervenciones cada vez menos intervencionistas necesita una mejor clasificación de estos pacientes. (13). Discernir la estabilidad de la columna es cada vez más necesario porque un conjunto de pacientes pueden ser tratados con técnicas de radioterapia complejas. En el marco de esta tesis se halla la reproducibilidad de la escala SINS entre los médicos que tratan esta patología y se compara con el diagnóstico establecido por el comité de tumores (12). Queda para un ulterior proyecto establecer su pronóstico futuro(14)

Las fracturas vertebrales no traumáticas son frecuentes en la práctica clínica. La mayoría son osteoporósticas (FVO) pero también son frecuentes las fracturas metastásicas (FVM). Su diagnóstico diferencial es fundamental para establecer un adecuada conducta diagnóstica y terapéutica. Se han propuesto varios hallazgos en imagen, más o menos asociados a un diagnóstico de referencia. Aunque se ha resaltado que su reproducibilidad debe comprobarse, son escasos los estudios sobre los mismos. Como la mejor técnica para el diagnóstico de las fracturas vertebrales es la RM, el propósito de esta tesis es triple (15): (a) la concordancia entre el diagnóstico basado en la imagen de FVM frente FVO y el patrón de referencia (biopsia o seguimiento mayor de 6 meses), (b) el acuerdo intra- e interobservador en los hallazgos claves de imagen y el diagnóstico de FVM vs FVO, y (c) evaluar si desvelar el dato de la historia del paciente respecto al cáncer lleva a variaciones en el diagnóstico, concordancia o acuerdo.

El acuerdo interobservador en medicina no ha sido suficiente explorado, aunque es inherente al hecho de observar en cualquier disciplina (16,17). Habitualmente, para las variables categóricas se utiliza el estadístico kappa y para las continuas el coeficiente de correlación intraclase (CCI). Ambos parámetros tienen limitaciones conocidas (16,18,19). Incluso en la especialidad de anatomía patológica, donde se examina el tejido, existen desacuerdos en la forma de clasificar la respuesta tumoral al tratamiento (20). Aunque la palpación es una maniobra ampliamente utilizada en la exploración física de la columna, sólo recientemente se ha comprobado la fiabilidad de la palpación en la columna torácica. Por ejemplo, la rigidez ha demostrado un acuerdo global menor que el debido al azar. Sin embargo el acuerdo para el dolor es moderado y el entrenamiento no mejoró el acuerdo (21). El famoso signo de Lasègue presenta un k=0,20-0,47, menor que el establecido para el diagnóstico de hernia frente a protrusión discal en imagen -k=0,44-0,59- (22,23). El acuerdo interobservador esperable en la imagen del raquis degenerativo es moderado, independientemente de la formación o el centro (24,25).

Existen serias limitaciones de la medicina basada en pruebas (MBP, *evidence-based medicine*, *EBM*) (26), motivadas porque se confunde reproducibilidad y fiabilidad con acuerdo (19). Una de las limitaciones de la MBP es que las propias pruebas en las que se debe basar la MBP, son inconsistentes o no se ha demostrado la fiabilidad de las mismas (27). En el raquis neoplásico se presentan artículos o aplicaciones web que se denominan basados en MBP (28) y su reproducibilidad no se ha contrastado hasta ahora (11,15,29). En esta tesis el marco fundamental es establecer la reproducibilidad básica de escalas y signos en imagen utilizados en la práctica diaria.

Hipótesis de trabajo

La hipótesis de este trabajo es que los diversos especialistas médicos implicados en la evaluación y tratamiento del raquis oncológico pueden alcanzar acuerdos razonables al aplicar en la práctica clínica las principales escalas y signos radiológicos de la columna vertebral metastásica en series de casos clínicos reales.

Objetivos

Objetivo general: conocer el acuerdo (intra- e inter-observador, e intra e inter-especialidades médicas) de los principales marcadores de estadificación de metástasis vertebrales, compresión medular, inestabilidad y diagnóstico diferencial con fracturas osteoporóticas entre los distintos médicos que tratan la columna metastásica.

Objetivos específicos:

- La estadificación de las metástasis vertebrales, valorado de manera separada mediante dos escalas distintas; Tomita y de Bauer.
- La existencia de compresión medular, valorada mediante la versión española de la "*Spinal cord compression scale*" (ESCC)
- La existencia de inestabilidad metastásica, valorada mediante la versión española del escala de inestabilidad de columna vertebral neoplásica (SINS) y comparación con el patrón de referencia del comité de tumores local.
- En las fracturas vertebrales acuerdo en diagnóstico debido a osteoporosis (FOV) o debido a metástasis vertebral (FMV), en los principales signos radiológicos y determinar si el hecho de que el especialista que interpreta la imagen conozca los eventuales antecedentes oncológicos del paciente, modifica su diagnóstico.

Conclusiones:

- El acuerdo en la estadificación de la enfermedad metastásica es alto con las escalas de Tomita y Bauer. Este sistema de escalas puede mejorar la comunicación entre médicos involucrados en la asistencia oncológica
- La escala ESCC puede ayudar a mejorar la comunicación de los médicos involucrados en la asistencia oncológica.

- El acuerdo en el estudio de la inestabilidad metastásica de la columna vertebral es moderado. La escala SINS puede mejorar la comunicación entre los médicos de asistencia oncológica.
- Con la RM para diagnosticar FVM frente a FVO, el acuerdo interobservador y la concordancia con el patrón de referencia fueron moderados. Estos resultados muestran las dudas en la reproducibilidad de basar estos diagnósticos en la práctica clínica con la RM.

Referencias

1. Coleman RE. Clinical features of metastatic bone disease and risk of skeletal morbidity. Clin Cancer Res. 2006 Oct 15;12(20 Pt 2):6243s-6249s.
2. Yao A, Sarkiss CA, Ladner TR, Jenkins AL. Contemporary spinal oncology treatment paradigms and outcomes for metastatic tumors to the spine: A systematic review of breast, prostate, renal, and lung metastases. J Clin Neurosci. 2017 Jul;41:11–23.
3. Wibmer C, Leithner A, Hofmann G, Clar H, Kapitan M, Berghold A, et al. Survival analysis of 254 patients after manifestation of spinal metastases: evaluation of seven preoperative scoring systems. Spine (Phila Pa 1976). 2011 Nov 1;36(23):1977–86.
4. Aoude A, Amiot L-P. A comparison of the modified Tokuhashi and Tomita scores in determining prognosis for patients afflicted with spinal metastasis. Can J Surg. 2014 Jun 1;57(3):188–93.
5. Ahmed AK, Goodwin CR, Heravi A, Kim R, Abu-Bonsrah N, Sankey E, et al. Predicting survival for metastatic spine disease: a comparison of nine scoring systems. Spine J. 2018 Oct;18(10):1804–14.
6. Karhade A V., Shin JH, Schwab JH. Prognostic models for spinal metastatic disease: evolution of methodologies, limitations, and future opportunities. Ann Transl Med. 2019 May;7(10):219–219.
7. Wang F, Zhang H, Yang L, Yang X, Zhang H, Li J, et al. Epidemiological Characteristics of 1196 Patients with Spinal Metastases: A Retrospective Study. Orthop Surg. 2019 Dec 22;11(6):1048–53.
8. Hui D, Mori M, Watanabe SM, Caraceni A, Strasser F, Saarto T, et al. Referral criteria for outpatient specialty palliative cancer care: an international consensus. Lancet Oncol. 2016 Dec 1;17(12):e552–9.
9. Bilsky MH, Laufer I, Fourney DR, Groff M, Schmidt MH, Varga PP, et al. Reliability analysis of the epidural spinal cord compression scale. J Neurosurg Spine. 2010 Sep;13(3):324–8.
10. Boussios S, Cooke D, Hayward C, Kanellos FS, Tsiouris AK, Chatziantoniou AA, et al. Metastatic Spinal Cord Compression: Unraveling the Diagnostic and Therapeutic Challenges. Anticancer Res. 2018 Sep 7;38(9):4987–97.
11. Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J, et al. Agreement in Metastatic Spinal Cord Compression. J Natl Compr Cancer Netw. 2016 Jan 1;14(1):70–6.
12. Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J, et al. Spine Instability Neoplastic Score: agreement across different medical and surgical specialties. Spine J. 2016 May;16(5):591–9.
13. Laufer I, Bilsky MH. Advances in the treatment of metastatic spine tumors: the future is not what it used to be. J Neurosurg Spine. 2019 Mar;30(3):299–307.
14. Donnellan CJ, Roser S, Maharaj MM, Davies BN, Ferch R, Hansen MA. Outcomes for Vertebrectomy for Malignancy and Correlation to the Spine Instability Neoplastic Score (SINS): a 10-Year Single-Center Perspective. World Neurosurg. 2020 Feb;
15. Arana E, Kovacs FM, Royuela A, Asenjo B, Nagib F, Pérez-Aguilera S, et al. Metastatic Versus Osteoporotic Vertebral Fractures on MRI: A Blinded, Multicenter, and Multispecialty Observer Agreement Evaluation. J Natl Compr Cancer Netw. 2020 Mar;18(3):267–73.
16. Schieda N. Interobserver Agreement of PI-RADS v. 2: Not All Features or Observers Are Created Equal. J Magn Reson Imaging. 2020 Feb;51(2):605–6.
17. Knottnerus JA, Tugwell P. Comprehensive evidence and clinical judgment. J Clin

Epidemiol. 2019 Aug;112:v–vi.
18. Shankar V, Bangdiwala SI. Observer agreement paradoxes in 2x2 tables: comparison of agreement measures. BMC Med Res Methodol. 2014 Aug 28;14:100.
19. Kottner J, Streiner DL. The difference between reliability and agreement. J Clin Epidemiol. 2011 Jun 1;64(6):701–2.
20. Verbeke C, Häberle L, Lenggenhager D, Esposito I. Pathology assessment of pancreatic cancer following neoadjuvant treatment: Time to move on. Pancreatology. 2018 Jul;18(5):467–76.
21. Beynon AM, Hebert JJ, Walker BF. The interrater reliability of static palpation of the thoracic spine for eliciting tenderness and stiffness to test for a manipulable lesion. Chiropr Man Therap. 2018 Dec 4;26(1):49.
22. Poiraudeau S, Foltz V, Drapé JL, Fermanian J, Lefèvre-Colau MM, Mayoux-Benhamou MA, et al. Value of the bell test and the hyperextension test for diagnosis in sciatica associated with disc herniation: comparison with Lasègue's sign and the crossed Lasègue's sign. Rheumatology (Oxford). 2001 Apr 1;40(4):460–6.
23. Jarvik JG, Deyo RA. Moderate versus mediocre: the reliability of spine MR data interpretations. Radiology. 2009 Jan 1;250(1):15–7.
24. Carrino JA, Lurie JD, Tosteson ANA, Tosteson TD, Carragee EJ, Kaiser J, et al. Lumbar spine: reliability of MR imaging findings. Radiology. 2009 Jan 1;250(1):161–70.
25. Arana E, Royuela A, Kovacs FM, Estremera A, Sarasíbar H, Amengual G, et al. Lumbar Spine: Agreement in the Interpretation of 1.5-T MR Images by Using the Nordic Modic Consensus Group Classification Form. Radiology. 2010 Mar;254(3):809–17.
26. Fava GA. Evidence-based medicine was bound to fail: a report to Alvan Feinstein. J Clin Epidemiol. 2017 Apr;84:3–7.
27. Feinstein AR, Horwitz RI. Problems in the "evidence" of "evidence-based medicine." Am J Med. 1997 Dec 1;103(6):529–35.
28. Wang KC, Jeanmenne A, Weber GM, Thawait SK, Thawait S, Carrino JA. An online evidence-based decision support system for distinguishing benign from malignant vertebral compression fractures by magnetic resonance imaging feature analysis. J Digit Imaging. 2011 Jun;24(3):507–15.
29. Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez U, Zamora J, et al. Agreement in the assessment of metastatic spine disease using scoring systems. Radiother Oncol. 2015 Apr;115(1):135–40.

Publicaciones originales que conforman la Tesis Doctoral con resumen

1. Acuerdo en el estudio de enfermedad metastásica de la columna vertebral con sistemas de puntuación

Radiotherapy and Oncology 115 (2015) 135–140

Objetivo: Comprobar la variabilidad de las escalas de Tomita y Bauer modificado en las metástasis vertebrales.

Material y Métodos: Los datos de imagen y clínicos de 90 pacientes con metástasis demostradas por biopsia se les proporcionaron a 83 especialista de 44 hospitales. Se determinaron dos veces por cada especialista los niveles afectados y las escalas de Tomita y Bauer modificados, con un intervalo mínimo de 6 semanas. Los médicos estaban ciegos a cada evaluación. Se empleó el estadístico kappa para comprobar el acuerdo intra- e interobservador. Los análisis de subgrupos se realizaron según la especialidad clínica (oncología médica, neurocirugía, radiología, cirugía ortopédica y oncología radioterápica), años de experiencia (≤ 7, 8-13, ≥ 14), y tipo de hospital (cuatro niveles).

Resultados: Para la identificación de las metástasis, el acuerdo intra-observador fue sustancial ($0,60<k<0,8$) en sacro y casi perfecto ($k>0,80$) en los otros niveles. El acuerdo interobservador fue casi perfecto en la columna lumbar y sustancial en los otros niveles. Para las escalas de Tomita y Bauer modificado el acuerdo intra-observador fue casi perfecto. El acuerdo interobservador fue casi perfecto para la escala de Tomita y sustancial para Bauer. Los resultados fueron similares entre especialidades, años de experiencia y tipo de hospital.

Conclusión: El acuerdo en el estudio de la enfermedad metastásica es alto. Este sistema de escalas puede mejorar la comunicación entre médicos involucrados en la asistencia oncológica.

Radiotherapy and Oncology 115 (2015) 135–140

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Metastatic spine disease

Agreement in the assessment of metastatic spine disease using scoring systems

Estanislao Arana [a,b,c,*], Francisco M. Kovacs [c,d], Ana Royuela [c,e,f], Beatriz Asenjo [c,g], Ursula Pérez-Ramírez [c,h], Javier Zamora [c,e,f,i], and the Spanish Back Pain Research Network Task Force for the improvement of inter-disciplinary management of spinal metastasis

[a] *Department of Radiology, Valencian Oncology Institute Foundation, Valencia;* [b] *Research Institute in Health Services Foundation, Valencia;* [c] *Spanish Back Pain Research Network, Kovacs Foundation;* [d] *Scientific Department, Kovacs Foundation, Palma de Mallorca;* [e] *CIBER Epidemiology and Public Health (CIBERESP);* [f] *Clinical Biostatistics Unit, Hospital Ramón y Cajal, IRYCIS, Madrid;* [g] *Department of Radiology, Hospital Regional Universitario Carlos Haya, Málaga;* [h] *Center for Biomaterials and Tissue Engineering, Universitat Politècnica de València, Valencia, Spain;* [i] *Barts and the London School of Medicine & Dentistry. Queen Mary University of London, London, UK*

ARTICLE INFO

Article history:
Received 15 January 2015
Received in revised form 13 March 2015
Accepted 15 March 2015
Available online 10 April 2015

Keywords:
Spine metastases
Intraobserver variability
Interobserver variability
Magnetic resonance imaging
Classification
Agreement

ABSTRACT

Purpose: To assess variability in the use of Tomita and modified Bauer scores in spine metastases.
Materials and methods: Clinical data and imaging from 90 patients with biopsy-proven spinal metastases, were provided to 83 specialists from 44 hospitals. Spinal levels involved and the Tomita and modified Bauer scores for each case were determined twice by each clinician, with a minimum of 6-week interval. Clinicians were blinded to every evaluation. Kappa statistic was used to assess intra and inter-observer agreement. Subgroup analyses were performed according to clinicians' specialty (medical oncology, neurosurgery, radiology, orthopedic surgery and radiation oncology), years of experience (⩽7, 8–13, ⩾14), and type of hospital (four levels).
Results: For metastases identification, intra-observer agreement was "substantial" ($0.60 < k < 0.80$) at sacrum, and "almost perfect" ($k > 0.80$) at the other levels. Inter-observer agreement was "almost perfect" at lumbar spine, and "substantial" at the other levels. Intra-observer agreement for the Tomita and Bauer scores was almost perfect. Inter-observer agreement was almost perfect for the Tomita score and substantial for the Bauer one. Results were similar across specialties, years of experience and type of hospital.
Conclusion: Agreement in the assessment of metastatic spine disease is high. These scoring systems can improve communication among clinicians involved in oncology care.

The organ most commonly affected by metastatic cancer is the skeleton, which is also where it causes the highest morbidity [1]. While new chemotherapeutic agents hinder the progression of spinal metastases and surgery is helpful for selected patients [2], radiotherapy continues to be the cornerstone in the treatment [3,4]. The Tomita and the modified Bauer scoring systems (Appendix 1) have been advocated as two of the most accurate methods for establishing the prognosis of metastatic spine disease and helping to select the most appropriate treatment for each case [5,6]. They are based on clinical data and imaging findings. However, few studies have analyzed the reliability of these scoring systems across different medical specialties [7], and none have assessed their intra and inter-observer agreement [8].

It has been reported that up to 98% of oncologists do not use a standardized method to assess the risk of pathological fracture [9]. Gathering data on the reliability of the Tomita and modified Bauer scores in clinical practice might be useful to promote their use when appropriate. Assessing the agreement among the different specialists involved in the assessment of spine metastatic disease (medical oncologists, radiologists, radiation oncologists, orthopedic surgeons and neurosurgeons), may contribute to improving consensus in the decision making process when determining the most suitable treatment for each patient.

Therefore, the purpose of this study was to assess the intra- and inter-observer agreement in the identification of the spine levels affected by metastatic cancer and in the calculation of the Tomita and modified Bauer scores, among a large sample of

* Corresponding author at: Fundación Instituto Valenciano de Oncología, C/Beltrán Báguena, 19, 46009 Valencia, Spain.
E-mail address: aranae@uv.es (E. Arana).

http://dx.doi.org/10.1016/j.radonc.2015.03.016

clinicians from different specialties, with varied degrees of experience and working in different settings and locations.

Methods

Study design and participants

This prospective study was approved by the institutional review boards of the participating hospitals, and complied with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [10].

Selection of hospital departments and clinicians

Sample size was calculated assuming that the prevalence of metastatic disease in a particular spinal level would be 10%, and that at least five readers would be recruited per medical specialty and five per hospital category; any increases in this prevalence or in the number of readers would therefore increase the power of this study. In order to ensure that, should kappa values reflect an "almost perfect" agreement ($k = 0.81$), the lower limit of the 95% confidence interval would lie within the range classified as reflecting a "substantial agreement" ($k = 0.61–0.80$), sample size was established at 90 patients. Sample size calculations where performed using kappa size package of the R library [11].

At the design phase of this study, the following medical specialties were defined as relevant for the clinical and therapeutic management of metastatic spine disease: neurosurgery, medical oncology, radiation oncology, radiology and orthopedic surgery.

All of the 132 physicians specialized in these clinical areas who had previously participated in studies undertaken by the Spanish Back Pain Research Network or had expressed interest in doing so, were invited to participate in this study. They worked in 61 hospital departments located in 44 hospitals across 14 out of the 17 Spanish regions; 12 were located in six private hospitals and the other 49 in 38 not for profit Hospitals, belonging to or working for the Spanish National Health Service (SNHS). The SNHS is the tax-funded, government run, organization which provides free health care to every resident in Spain.

The SNHS classifies Hospitals in five types, based on the size of the catchment area, number of beds, number of clinicians, availability of high tech medical equipment and procedures, education, training and academic activity, and clinical complexity of the cases treated (i.e., being the "reference hospital" for specific diseases or procedures) [12]. Category 1 is the simplest and category 5 is the most complex.

Specialists invited to participate in this study as readers worked in hospitals belonging to categories 2, 3, 4 and 5. Those who accepted were asked to provide the number of years they had been working in clinical practice after their residency. According to the usual policy within the SNHS, the Departments and clinicians did not receive any compensation for participating in this study.

Patients and images selection

Patients and images were selected by an oncoradiologist who worked in a category 4 hospital and did not act as a reader in this study. He reviewed consecutive MRIs performed in his Radiology Department on patients who had been diagnosed as presenting spine metastatic disease by the tumor boards at his hospital. These cases were revised in inverse chronological order (i.e., MRIs performed more recently were revised first). All exams had been acquired on the same 1.5 T superconducting system with a phased-array multicoil (Siemens Symphony, Erlangen, Germany), in the supine position with a fixed imaging protocol. The radiologist selected the two most representative sagittal images per patient; one T1 weighted image and one short tau inversion-recovery (STIR) weighted image [13].

The first 90 cases which complied with inclusion criterion and not with exclusion criteria, were selected. Inclusion criterion was presenting a stage IV (AJCC classification 7th Edition, 2010) spine metastatic disease confirmed by biopsy from the primary tumor site and from one of the spine metastases. Exclusion criteria were; (a) clinical history lacking data required to assess Tomita and modified Bauer scores, or (b) imaging of insufficient quality to assess the spinal level/s affected.

Procedure

The recruiting radiologist prepared an information pack corresponding to each patient, comprising the two MR images and a clinical vignette which included patient's age, oncologic history, clinical signs and symptoms (Supplementary Fig. S1) [14]. Patient identity was masked and a code was assigned to each information pack. All the information packs were uploaded to an online platform specifically designed for this study (http://www.typeform.com/).

Each reader was provided with a personal password to assess the information packs online. Readers were asked to indicate all the spinal levels in which they detected metastases for each patient (cervical, thoracic, lumbar, and/or sacral). They were only provided with definitions included in the Tomita and modified Bauer scores, as shown in Table 1. Visceral metastases included in this study were not treatable with surgery or focal therapies; therefore, they were considered as non-treatable when calculating the Tomita score [5]. No attempt was made to further define or standardize the meaning of the terms included in the scoring systems or to homogenize the diagnostic criteria, and readers did not receive any instructions regarding the interpretation of images. They were told to use their own clinical judgment when in doubt, as they would do in every-day, routine clinical practice.

Readers assessed the information pack alone and on their own, and uploaded the resulting report directly onto the online platform. They were asked to assess the same clinical sets twice, with a minimum of six weeks' interval. The platform software ensured that the minimum period was observed, and that readers had no access to their own previous reports or to their colleagues' current or previous reports.

All reports were entered into the database at a centralized coordination office. Data introduced into the platform were automatically converted into a spreadsheet. The software engineer in charge of developing the platform crosschecked that data in the database matched the information introduced into the platform by the readers.

Statistical analysis

The scores on the Tomita and modified Bauer scoring systems were grouped according to the treatment they imply. Therefore, scores on the modified Bauer scoring system were classified into three categories; 0–1 (supportive care); 2 (short term palliation), and 3–4 (middle term local control) [7]; while Tomita scores were classified into four categories; 2–3 (long-term local control); 4–5 (middle-term local control); 6–7 (short-term palliation), and 8–10 (terminal care) [5]. Data on the presence of metastases at each spinal level in each patient was classified as yes or no.

To assess intra- and inter-observer agreement, ratings from each reader were cross-tabulated and the kappa statistic was calculated. A weighted-kappa approach, with bi-squared weighting scheme, was used for the analysis of the agreement when using the Tomita and modified Bauer scoring systems. Kappa values were categorized as "almost perfect" (0.81–1.00), "substantial" (0.61–0.80), "moderate" (0.41–0.60), "fair" (0.21–0.40), "slight" (0.00–0.20), and "poor" (<0.00) [15].

Table 1
Sample characteristics.

Hospitals[a]		44
	Degree of complexity[b]	
	Category 2	3 (6.8)
	Category 3	11 (25)
	Category 4	9 (20.4)
	Category 5	21 (47.7)
	Management	
	Not for profit	38
	For profit	6
Departments[a]		61
	Radiology	19 (31.1)
	Radiation oncology	11 (18.0)
	Orthopedic surgery	12 (19.7)
	Neurosurgery	12 (19.7)
	Medical oncology	7 (11.5)
Readers[a]		83
	Specialty	
	Radiology	23 (27.7)
	Radiation oncology	22 (26.5)
	Orthopedic surgery	16 (19.3)
	Neurosurgery	14 (16.9)
	Medical oncology	8 (9.6)
	Years in practice (post-residency)	
	⩽7	27 (32.5)
	8–13	25 (30.1)
	⩾14	31 (37.4)
	Setting	
	Category of hospital in which they work	
	Category 2	3 (3.6)
	Category 3	25 (30.1)
	Category 4	19 (22.9)
	Category 5	36 (43.4)
	Hospital management[c]	
	Not for profit	71
	For profit	12
Patients		90
	Age (years)[d]	60.8 (12.3)
	Gender (males)[a]	39 (43.3)
	Cancer subtype[a]	
	Breast carcinoma	37 (41.1)
	Prostate carcinoma	16 (17.8)
	Lung carcinoma	12 (13.3)
	Renal cell carcinoma	6 (6.7)
	Endometrial carcinoma	3 (3.3)
	Unknown origin	2 (2.2)
	Small bowel carcinoma	2 (2.2)
	Melanoma	2 (2.2)
	Ovarian carcinoma	2 (2.2)
	Hemangiosarcoma	2 (2.2)
	Ewing sarcoma	1 (1.1)
	Testicular germ cell tumor	1 (1.1)
	Cervix carcinoma	1 (1.1)
	Gastric carcinoma	1 (1.1)
	Hepatocellular carcinoma	1 (1.1)
	Urinary bladder carcinoma	1 (1.1)
	Location of metastases[a]	
	Cervical	4 (4.4)
	Cervical and thoracic	15 (16.7)
	Cervical, thoracic and lumbar	1(1.1)
	Cervical, thoracic, lumbar and sacral	2 (2.2)
	Thoracic	18 (20)
	Thoracic and lumbar	15 (16.7)
	Thoracic, lumbar and sacral	24 (26.7)
	Lumbar	5 (5.6)
	Lumbar and sacral	6 (6.7)
Spinal levels involved[a]		182
	Cervical	22 (12.1)
	Thoracic	75 (41.2)
	Lumbar	53 (29.1)
	Sacral	32 (17.6)

[a] *n* (%).
[b] Category of hospital; complexity (based on size, availability of high tech medical equipment and procedures, education activity, etc.) ranges from category 1 (the simplest-none of this type were included in this study) to category 5 (the most complex). See text for details.
[c] Not for profit: Hospitals belonging to the Spanish National Health Service (SNHS) or to charities working for the SNHS. For profit: hospitals under private ownership and management.
[d] Mean (SD).

To assess intra-observer agreement for each variable (Tomita score, modified Bauer score, and level/s involved), a kappa index was calculated for each one of the 83 readers, and median, 5th and 95th percentiles values were calculated.

To assess inter-observer agreement, the corresponding kappa index was calculated and the 95% Confidence Interval (95% CI) was determined following the jackknife resampling method [16].

Subgroup analyses for each variable were performed, in which kappa values were calculated separately depending on medical specialty, hospital category and professional experience. Degree of professional experience was classified as "recently specialized" (⩽7 years in practice, after residency), "experienced" (8–13 years), and "senior specialist" (⩾14 years).

Statistical package Stata v 13. (StataCorp. 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP) was used.

Results

Eighty-three (62.87%) out of the 132 clinicians who were invited to act as readers, participated in this study; 23 radiologists, 22 radiation oncologists, 16 orthopedic surgeons, 14 neurosurgeons, and 8 medical oncologists, working in 61 hospital departments.

The first 90 patients selected by the recruiting radiologist (51 women and 39 men, mean age 60.8 years) complied with the inclusion criteria, and none was excluded. The number of spinal levels involved was 182 Table 1 shows sample characteristics.

There were more than five readers for each specialty and degree of professional experience. However, only three readers worked at category 2 hospitals; therefore, agreement for this subgroup was not calculated (Appendices 2 and 3, Tables 2 and 3).

Intra-observer agreement in the identification of the spinal levels involved was "almost perfect" except for those located at the sacral level, for which it was "substantial" (Appendix 2). Subgroup analyses showed that these results were unaffected by readers' degree of experience, and category of the hospital in which they worked. Intra-observer agreement among radiologist was "almost perfect" at all spinal levels (including sacral) and was "substantial" at the thoracic level for neurosurgeons and orthopedic surgeons (Appendix 2).

Inter-observer agreement in the identification of the affected spinal levels was "substantial" except at the lumbar level, for which it was "almost perfect" (Appendix 3). Subgroup analyses showed the following particularities; (1) agreement at the sacral level was "moderate" among orthopedic surgeons, radiation oncologists and readers working in category 5 hospitals; and "fair" among radiologists and readers with >14 years of experience. (2) Agreement at the thoracic level was "almost perfect" for neurosurgeons. (3) Agreement at the cervical level was "almost perfect" for orthopedic surgeons, and "moderate" for medical oncologists (Appendix 3).

Intra-observer agreement in the Tomita and modified Bauer scores was "almost perfect". In subgroup analyses, the only exception was that agreement among medical oncologists when using the modified Bauer score was "substantial" (Table 2).

Inter-observer agreements in the Tomita and modified Bauer scores were "almost perfect" and "substantial", respectively

Table 2
Intra-observer agreement among 83 readers in the modified Bauer and Tomita scores.[a]

	Bauer score	Tomita score
Global intra-observer agreement	0.884 (0.718; 0.973)	0.960 (0.814; 0.996)
By specialty		
Orthopedic surgery	0.872 (0.646; 0.973)	0.944 (0.471; 0.996)
Neurosurgery	0.889 (0.754; 0.973)	0.944 (0.845; 0.993)
Medical oncology	0.761 (0.387; 0.964)	0.919 (0.496; 0.978)
Radiation oncology	0.888 (0.828; 0.971)	0.966 (0.865; 0.996)
Radiology	0.907 (0.818; 0.991)	0.970 (0.854; 0.996)
By experience		
⩽7 years	0.884 (0.667; 0.955)	0.955 (0.496; 0.989)
8–13 years	0.883 (0.734; 0.973)	0.960 (0.873; 0.993)
⩾14 years	0.901 (0.734; 0.991)	0.962 (0.814; 0.996)
By hospital complexity[b]		
Category 2[c]	–	–
Category 3	0.880 (0.710; 0.957)	0.955 (0.496; 0.993)
Category 4	0.884 (0.646; 0.991)	0.964 (0.774; 0.996)
Category 5	0.890 (0.734; 0.979)	0.958 (0.858; 0.996)

[a] κ values: median (5th; 95th percentiles).
[b] Complexity (based on size, availability of high tech medical equipment and procedures, education activity, etc.) ranges from category 1 (the simplest -none of this category were included in this study) to category 5 (the most complex). See text for details.
[c] Only three specialists working in category 2 hospitals participated in this study. Therefore, agreement was not calculated for this subgroup.

Table 3
Inter-observer agreement among 83 readers in the modified Bauer and Tomita scores.[a]

	Bauer score	Tomita score
Global inter-observer agreement	0.790 (0.746; 0.840)	0.905 (0.881; 0.932)
By specialty		
Orthopedic surgery	0.786 (0.679; 0.933)	0.893 (0.823; 0.986)
Neurosurgery	0.769 (0.627; 0.958)	0.863 (0.765; 0.999)
Medical oncology	0.732 (0.409; 1.000)	0.933 (0.866; 1.000)
Radiation oncology	0.771 (0.677; 0.894)	0.894 (0.839; 0.963)
Radiology	0.797 (0.721; 0.898)	0.914 (0.879; 0.961)
By experience		
⩽7 years	0.774 (0.687; 0.884)	0.873 (0.811; 0.951)
8–13 years	0.779 (0.687; 0.896)	0.910 (0.867; 0.964)
⩾14 years	0.811 (0.748; 0.890)	0.916 (0.886; 0.954)
By category of hospital[b]		
Category 2[c]	–	–
Category 3	0.758 (0.665; 0.877)	0.863 (0.795; 0.949)
Category 4	0.784 (0.665; 0.939)	0.903 (0.838; 0.988)
Category 5	0.793 (0.728; 0.874)	0.912 (0.883; 0.948)

[a] κ values; median (95% CI).
[b] Complexity (based on size, availability of high tech medical equipment and procedures, education activity, etc.) ranges from category 1 (the simplest -none of this category were included in this study) to category 5 (the most complex). See text for details.
[c] Only three specialists working in category 2 hospitals participated in this study. Therefore, agreement was not calculated for this subgroup.

(Table 3). Subgroup analyses showed that agreement in the use of the Tomita score was higher than the one derived from the modified Bauer score irrespective of specialty, degree of experience and hospital category. When using the Tomita score, the agreement was "almost perfect" across all categories, while the agreement in the use of the modified Bauer score was "almost perfect" only among readers with ⩾14 years of experience (Table 3).

Discussion

This study did not implement any measures for improving inter-observer agreement (such as agreeing on diagnostic criteria or using available online examples linked to standardized nomenclature) [17,18]. A high number of readers participated, they had different backgrounds and worked in different hospitals located in different regions; most readers had never met their colleagues in person. Nevertheless, results from this study reflect a high degree of agreement among clinicians involved in the management of spine metastatic disease, both when identifying the spinal levels affected and when using the scoring systems which help to establish a prognosis and determine the appropriate treatment. In the subgroup analyses, the classification of the degree of agreement varied across some subgroups (i.e., "almost perfect" vs. "substantial"), but actual differences in kappa values were small and likely to be clinically irrelevant [10]. This is reassuring for patients, since it suggests that the decision-making process is reasonably consistent irrespective of differences in the specialty of the treating clinician, number of years of post-residency clinical practice, and hospital characteristics.

Very few studies have focused on the agreement in the assessment of metastatic spine diseases across specialties. A previous one found a "moderate" to "poor" agreement between musculoskeletal radiologists and orthopedic surgeons when assessing certain imaging features of spinal metastases [8]. Differences in the specialties being compared may account for differences in results from the current study.

Bone marrow abnormalities are common in the sacral area [19], which may make it more challenging to detect metastases [20,21]. This may account for the fact that, in this study, the finding with the lowest intra- and inter-observer agreements was the presence of metastases at the sacral level (Appendices 2 and 3).

The assessment of imaging by spine surgeons is usually considered as the gold standard for deciding whether surgery is appropriate for a patient with metastatic spine disease [22]. However, the degree of agreement found in this study was similar across specialties and hospital categories (i.e., irrespective of their size and complexity) (Appendices 2 and 3, Tables 2 and 3). This suggests that the scoring systems are intrinsically reliable, and may also reflect appropriate training standards, continuous medical education, a solid pattern of common knowledge, and good interdisciplinary communication among clinicians managing oncologic patients [4]. These results are reassuring from the point of view of equity in health care, and are generally consistent with previous studies which have found no differences between conventional hospitalization and quick diagnosis units [23].

Clinicians involved in the management of oncology patients come from a variety of backgrounds. Good communication among them leads to consistency of care, which is a prerequisite for effectiveness [24]. Results from this study suggest that the use of the modified Bauer or Tomita scoring systems can be useful for accurate communication among multidisciplinary team members [2].

Previous studies have shown that the modified Bauer scoring system is simpler (Appendix 1) and predicts survival better than the Tomita one [5,6]. The agreement found in this study was higher for the latter; values obtained when using the Tomita score are consistently higher in all the analyses, for intra and inter-observer agreement, for the whole sample and for all the subgroups (Tables 2 and 3). However differences in the kappa values are small, so they should not be seen as the key criterion for selecting the scoring system to be used in the clinical environment [2,10].

Some studies have found that an improvement in the quality of care for oncologic patients does not necessarily translate immediately into better clinical results or improved survival rates [25]. Similarly, it may be argued that "agreement" when using a scoring system, does not necessarily mean that the resulting recommended treatment is "appropriate" or that it will lead to improved survival rates [2]. In fact, it is theoretically possible for clinicians to agree on measures which are not evidence-based or effective [26] and even reluctance to practice evidence-based radiotherapy [3]. Moreover, some features of the modified Bauer and Tomita scoring systems may be criticized; for instance, while it is widely accepted that the origin of the primary tumor is the most important prognostic factor for survival [5,7], these scoring systems do not explicitly consider all subtypes of cancer which can cause spinal metastases [2,27]. However, this study did not focus on assessing the validity of the Tomita or modified Bauer scoring systems as a tool for identifying the most effective treatment for each patient, or on measuring the improvements in clinical results generated by their use, as other prognostic factors may be valid [27,28]. This study focused on assessing the degree of agreement when using these scoring systems. A high degree of intra- and inter-observer agreement (Tables 2 and 3), suggests that these instruments are reliable and can be used for inter-disciplinary communication among the clinicians involved in each case. This supports their use in clinical practice, but also in trials intended to assess whether they improve clinical results, and in studies aiming to fine-tune therapeutic decisions based on their score, as radiotherapy for non-surgical candidates [2,4].

This study has some potential limitations. Readers were only provided with two selected images per case. It is possible that supplying every reader with all the images available for each patient might have changed the degree of agreement. However, providing the readers with a selection of images ensures that all assess the same ones, and fits in with the procedure usually followed to appraise reliability [14]. Physicians who acted as readers in this study were volunteers, and were not randomly selected. Therefore, selection bias may exist; it is possible that physicians who agreed to participate in this study were the most motivated or interested in spine metastatic disease [29]. Should this be the case, agreement might have been lower had other clinicians less familiar with spine metastatic disease, participated. It is impossible to completely rule out this possibility. However, this does not question the intrinsic reliability of the scoring systems used. Moreover, the degree of agreement was high, despite the fact that no measures were implemented to facilitate agreement, that the number of participants was large and that they came from different specialties and settings. All of the above suggests that their use in routine clinical practice is reliable.

In conclusion, this study suggests that the agreement of radiologists, medical oncologists, radiation oncologists, orthopedic surgeons and neurosurgeons when identifying the spinal level affected by metastases and using the Tomita and modified Bauer scoring systems, is high.

Conflicts of interest

No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this article.

The authors do not have any financial or personal relationships with third parties that could influence this work inappropriately. The authors have no conflicts of interest to report.

Funding

This study was funded by the Kovacs Foundation, a not-for-profit Spanish institution specializing in neck and back pain research, and with no links to the health industry.

Acknowledgements

Thank you to Prof David Moratal for his collaboration.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.radonc.2015.03.016.

References

[1] Coleman RE. Clinical features of metastatic bone disease and risk of skeletal morbidity. Clin Cancer Res 2006;12:6243s–9s. http://dx.doi.org/10.1158/1078-0432.CCR-06-0931.
[2] Kaloostian PE, Yurter A, Zadnik PL, Sciubba DM, Gokaslan ZL. Current paradigms for metastatic spinal disease: an evidence-based review. Ann Surg Oncol 2014;21:248–62. http://dx.doi.org/10.1245/s10434-013-3324-8.
[3] Popovic M, den Hartogh M, Zhang L, Poon M, Lam H, Bedard G, et al. Review of international patterns of practice for the treatment of painful bone metastases with palliative radiotherapy from 1993 to 2013. Radiother Oncol 2014;111:11–7. http://dx.doi.org/10.1016/j.radonc.2014.01.015.
[4] Lutz ST, Jones J, Chow E. Role of radiation therapy in palliative care of the patient with cancer. J Clin Oncol 2014;32:2913–9. http://dx.doi.org/10.1200/JCO.2014.55.1143.
[5] Tomita K, Kawahara N, Kobayashi T, Yoshida A, Murakami H, Akamaru T. Surgical strategy for spinal metastases. Spine (Phila Pa 1976) 2001;26:298–306. doi: 10.1097/00007632-200102010-00016.
[6] Wibmer C, Leithner A, Hofmann G, Clar H, Kapitan M, Berghold A, et al. Survival analysis of 254 patients after manifestation of spinal metastases: evaluation of seven preoperative scoring systems. Spine (Phila Pa 1976) 2011;36:1977–86. doi: 10.1097/BRS.0b013e3182011f84.
[7] Leithner A, Radl R, Gruber G, Hochegger M, Leithner K, Welkerling H, et al. Predictive value of seven preoperative prognostic scoring systems for spinal metastases. Eur Spine J 2008;17:1488–95. http://dx.doi.org/10.1007/s00586-008-0763-1.
[8] Khan L, Mitera G, Probyn L, Ford M, Christakis M, Finkelstein J, et al. Inter-rater reliability between musculoskeletal radiologists and orthopedic surgeons on computed tomography imaging features of spinal metastases. Curr Oncol 2011;18:e282–7. http://dx.doi.org/10.3747/co.v18i6.797.
[9] Cumming D, Cumming J, Vince A, Benson R. Metastatic bone disease: the requirement for improvement in a multidisciplinary approach. Int Orthop 2009;33:493–6. http://dx.doi.org/10.1007/s00264-008-0556-5.
[10] Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol 2011;64:96–106. http://dx.doi.org/10.1016/j.jclinepi.2010.03.002.
[11] Donner A, Rotondi MA. Sample size requirements for interval estimation of the kappa statistic for interobserver agreement studies with a binary outcome and multiple raters. Int J Biostat 2010;6. doi: 10.2202/1557-4679.1275.
[12] Departamento de Métodos Cuantitativos en Economía y Gestión U de LP de GC. Clasificación de hospitales públicos españoles mediante el uso del análisis cluster; 2007. http://www.icmbd.es/docs/resumenClusterHospitales.pdf [accessed 10.10.14].
[13] Shen G, Deng H, Hu S, Jia Z. Comparison of choline-PET/CT, MRI, SPECT, and bone scintigraphy in the diagnosis of bone metastases in patients with prostate cancer: a meta-analysis. Skeletal Radiol 2014;43:1503–13. http://dx.doi.org/10.1007/s00256-014-1903-9.
[14] Fourney DR, Frangou EM, Ryken TC, Dipaola CP, Shaffrey CI, Berven SH, et al. Spinal instability neoplastic score: an analysis of reliability and validity from the spine oncology study group. J Clin Oncol 2011;29:3072–7. http://dx.doi.org/10.1200/JCO.2010.34.3897.
[15] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159. http://dx.doi.org/10.2307/2529310.
[16] Efron B. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. Biometrika 1981;68:589–99. http://dx.doi.org/10.1093/biomet/68.3.589.
[17] Brorson S, Hróbjartsson A. Training improves agreement among doctors using the Neer system for proximal humeral fractures in a systematic review. J Clin Epidemiol 2008;61:7–16. http://dx.doi.org/10.1016/j.jclinepi.2007.04.014.
[18] Lee EH, Jun JK, Jung SE, Kim YM, Choi N. The efficacy of mammography boot cAMP to improve the performance of radiologists. Korean J Radiol 2014;15:578–85. http://dx.doi.org/10.3348/kjr.2014.15.5.578.
[19] Long SS, Yablon CM, Eisenberg RL. Bone marrow signal alteration in the spine and sacrum. AJR Am J Roentgenol 2010;195:W178–200. http://dx.doi.org/10.2214/AJR.10.4331.
[20] Sciubba DM, Petteys RJ, Garces-Ambrossi GL, Noggle JC, McGirt MJ, Wolinsky J-P, et al. Diagnosis and management of sacral tumors. J Neurosurg Spine 2009;10:244–56. http://dx.doi.org/10.3171/2008.12.SPINE08382.
[21] Chung HT, Lee B, Park E, Lu JJ, Xia P. Can all centers plan intensity-modulated radiotherapy (IMRT) effectively? An external audit of dosimetric comparisons between three-dimensional conformal radiotherapy and IMRT for adjuvant chemoradiation for gastric cancer. Int J Radiat Oncol Biol Phys 2008;71:1167–74. http://dx.doi.org/10.1016/j.ijrobp.2007.11.040.
[22] Fisher CG, Versteeg AL, Schouten R, Boriani S, Varga PP, Rhines LD, et al. Reliability of the spinal instability neoplastic scale among radiologists: an

assessment of instability secondary to spinal metastases. AJR Am J Roentgenol 2014;203:869–74. http://dx.doi.org/10.2214/AJR.13.12269.
[23] Bosch X, Moreno P, Ríos M, Jordán A, López-Soto A. Comparison of quick diagnosis units and conventional hospitalization for the diagnosis of cancer in Spain: a descriptive cohort study. Oncology 2012;83:283–91. http://dx.doi.org/10.1159/000341658.
[24] Fleissig A, Jenkins V, Catt S, Fallowfield L. Multidisciplinary teams in cancer care: are they effective in the UK? Lancet Oncol 2006;7:935–43. http://dx.doi.org/10.1016/S1470-2045(06)70940-8.
[25] Keating NL, Landrum MB, Lamont EB, Bozeman SR, Shulman LN, McNeil BJ. Tumor boards and the quality of cancer care. J Natl Cancer Inst 2013;105:113–21. http://dx.doi.org/10.1093/jnci/djs502.
[26] Dea N, Fisher CG. Evidence-based medicine in metastatic spine disease. Neurol Res 2014;36:524–9. http://dx.doi.org/10.1179/1743132814Y.0000000365.
[27] Bollen L, van der Linden YM, Pondaag W, Fiocco M, Pattynama BPM, Marijnen C, et al. Prognostic factors associated with survival in patients with symptomatic spinal bone metastases: a retrospective cohort study of 1,043 patients. Neuro Oncol 2014;16:991–8. http://dx.doi.org/10.1093/neuonc/not318.
[28] Krishnan MS, Epstein-Peterson Z, Chen Y-H, Tseng YD, Wright AA, Temel JS, et al. Predicting life expectancy in patients with metastatic cancer receiving palliative radiotherapy: the TEACHH model. Cancer 2014;120:134–41. http://dx.doi.org/10.1002/cncr.28408.
[29] Lim C, Cheung MC, Franco B, Dharmakulaseelan L, Chong E, Iyngarathasan A, et al. Quality improvement: an assessment of participation and attitudes of medical oncologists. J Oncol Pract 2014;10:e408–14. http://dx.doi.org/10.1200/JOP.2014.001515.

2. Acuerdo en la compresión medular metastásica

J Natl Compr Canc Netw 2016;14:70–76

Antecedentes: la compresión medular metastásica epidural (ESCC, en su acrónimo inglés) es una urgencia médica gravísima. El objetivo de este estudio fue determinar la fiabilidad del sistema de escala de 6 puntos de ESCC y la identificación del nivel vertebral que presenta ESCC.

Métodos: Los datos de imagen y clínicos de 90 pacientes con metástasis demostradas por biopsia se les proporcionaron a 83 especialista de 44 hospitales. Cada médico determinó el nivel que presentaba las metástasis y la escala ESCC dos veces, con un intervalo mínimo de 6 semanas. Los médicos estaban ciegos a cada evaluación suya y a la de los otros médicos. El kappa de Fleiss (k) se usó para comprobar el acuerdo intra- e interobservador. Los análisis de subgrupos se realizaron según la especialidad clínica (oncología médica, neurocirugía, radiología, cirugía ortopédica y oncología radioterápica), años de experiencia (≤ 7, 8-13, ≥ 14), y tipo de hospital.

Resultados: El acuerdo intraobservador e interobservador en la localización de ESCC fue sustancia (k>0,61). El acuerdo intraobservador en la escala ESS fue excelente (k=0,82), mientras el acuerdo interobservador fue sustancial (k=0,64). El acuerdo global con la clasificación del comité tumoral fue sustancial (k=0,71). Los resultados fueron similares entre especialidades, años de experiencia y categoría hospitalaria

Conclusiones: La escala ESCC puede ayudar a mejorar la comunicación de los médicos involucrados en la asistencia oncológica.

Original Research

Agreement in Metastatic Spinal Cord Compression

Estanislao Arana, MD, MHE, PhD[a,b,c]; Francisco M. Kovacs, MD, PhD[c,d]; Ana Royuela, PhD[c,e,f]; Beatriz Asenjo, MD, PhD[c,g]; Úrsula Pérez-Ramírez, MSc[c,h]; Javier Zamora, PhD[c,e,f,i]; and the Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis*

Abstract

Background: Metastatic epidural spinal cord compression (ESCC) is a devastating medical emergency. The purpose of this study was to determine the reliability of the 6-point ESCC scoring system and the identification of the spinal level presenting ESCC. **Methods:** Clinical data and imaging from 90 patients with biopsy-proven spinal metastases were provided to 83 specialists from 44 hospitals. The spinal levels presenting metastases and the ESCC scores for each case were calculated twice by each clinician, with a minimum of 6 weeks' interval. Clinicians were blinded to assessments made by other specialists and their own previous assessment. Fleiss kappa (κ) statistic was used to assess intraobserver and interobserver agreement. Subgroup analyses were performed according to clinicians' specialty (medical oncology, neurosurgery, radiology, orthopedic surgery, and radiation oncology), years of experience, and type of hospital. **Results:** Intraobserver and interobserver agreement on the location of ESCC was substantial (κ>0.61). Intraobserver agreement on the ESCC score was "excellent" (κ=0.82), whereas interobserver agreement was substantial (κ=0.64). Overall agreement with the tumor board classification was substantial (κ=0.71). Results were similar across specialties, years of experience and hospital category. **Conclusions:** The ESCC score can help improve communication among clinicians involved in oncology care.

J Natl Compr Canc Netw 2016;14(1):70–76

Background

Malignant or metastatic spinal cord compression of the thecal sac is a devastating medical emergency presented by 5% to 20% of patients with spinal metastases.[1] It can be caused by vertebral collapse, but is usually provoked by soft tissue causing epidural spinal cord compression (ESCC).[2]

Clinical symptoms and the ESCC grade are the major determinants in the decision to operate or irradiate.[3,4] The ESCC score system is a 6-point scale for diagnosing and reporting ESCC based on imaging findings (Figure 1).[3] It was developed by oncologic spine surgeons and proven to be reliable among a small sample of these specialists.[3] However, managing spinal cord compression requires a multidisciplinary approach,[5,6] and the lack of nomenclature standardization prevents agreement in decision-making,[7] delays appropriate treatment, and hinders treatment effectiveness.[8,9]

Therefore, the purpose of this study was to assess intraobserver and interobserver agreement in identification of spine level involved in each patient and in the calculation of the ESCC score among a large sample of clinicians from different specialties with varied degrees of experience and working in different settings and locations.

Methods

This study was approved by the Institutional Review Boards of the participating hospitals, and complied with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).[10]

From the [a]Department of Radiology, Valencian Oncology Institute Foundation, Valencia, Spain; [b]Research Institute in Health Services Foundation, Valencia, Spain; [c]Spanish Back Pain Research Network, Kovacs Foundation, Palma de Mallorca, Spain; [d]Scientific Department, Kovacs Foundation, Palma de Mallorca, Spain; [e]CIBER Epidemiology and Public Health (CIBERESP), Madrid, Spain; [f]Clinical Biostatistics Unit, Hospital Ramón y Cajal, IRYCIS, Madrid, Spain; [g]Department of Radiology, Hospital Regional Universitario Carlos Haya, Málaga, Spain; [h]Center for Biomaterials and Tissue Engineering, Universitat Politècnica de València, Valencia, Spain; and [i]Barts and the London School of Medicine & Dentistry, Queen Mary University of London, London, United Kingdom.

*The view members of the Spanish Back Pain Research Network Task Force, see supplemental eAppendix 1 (available with this issue, at JNCCN.org).

Submitted June 22, 2015; accepted for publication September 4, 2015.

The authors have disclosed that they have no financial interests, arrangements, affiliations, or commercial interests with the manufacturers of any products discussed in this article or their competitors. Funding for this work was provided by The Kovacs Foundation.

Correspondence: Estanislao Arana, MD, MHE, PhD, Fundación Instituto Valenciano de Oncología, C/ Beltrán Báguena, 19, 46009 Valencia, Spain. E-mail: aranae@uv.es

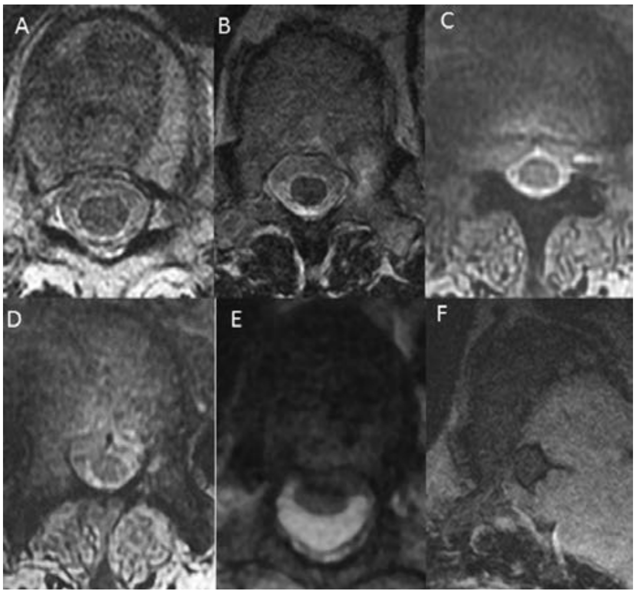


Figure 1. The 6-point ESCC classification according to Bilsky et al.[3] (A) Grade 0 indicates bone-only disease. (B) Grade 1a, epidural impingement, without deformation of the thecal sac. (C) Grade 1b, deformation of the thecal sac, without spinal cord abutment. (D) Grade 1c, deformation of the thecal sac with spinal cord abutment, but without cord compression. (E) Grade 2, spinal cord compression, but with cerebrospinal fluid (CSF) visible around the cord. (F) Grade 3, spinal cord compression, no CSF visible around the cord.

Selection of Hospital Departments and Clinicians

Sample size was calculated assuming that the prevalence of ESCC in a particular spinal level would be 10%, and that at least 5 readers would be recruited per medical specialty and 5 per hospital category. In order to ensure that, should κ values reflect an "almost perfect" agreement (κ=0.81), the the lower limit of the 95% CI would lie within the range classified as reflecting a substantial agreement (κ=0.61–0.80), the sample size was established at 90 patients. Sample size calculations were performed using κ size package of the R library.[11]

The 61 hospital departments specializing in radiology, medical oncology, radiation oncology, neurosurgery, and orthopedic surgery, which had previously participated in studies undertaken by the Spanish Back Pain Research Network (REIDE) or had expressed interest in doing so, were invited to participate in this study. Twelve departments were located in 6 private hospitals and the other 49 in 38 nonprofit hospitals, belonging to or working for the Spanish National Health Service (SNHS). The SNHS is the tax-funded, government-owned organization that provides free health care to every resident in Spain.

The SNHS classifies hospitals into 5 categories based on the size of the catchment area; number of beds; number of clinicians; availability of high-tech medical equipment and procedures; education, training, and academic activity; and clinical complexity of the cases treated (ie, being the reference hospital for specific diseases or procedures).[12] Category 1 is the simplest and category 5 is the most complex. Departments invited to participate in this study were located in hospitals belonging to categories 2, 3, 4, and 5.

All clinicians who had finished their residency and worked at the participating departments were invited to act as readers in this study, and asked to report the number of years they had been working in clinical practice after their residency. The departments and clinicians did not receive any compensation for participating in this study.

Patients and Images Selection

A radiologist at a category 4 hospital, who did not act as reader, was responsible for selecting patients and images for study inclusion. He identified patients who had undergone an MRI in his department for spinal cord compression and whose ESCC scores had been rated by a tumor board.

The tumor board comprised a medical oncologist, a radiation oncologist, a radiologist, a pathologist, an orthopedic surgeon, and a neurosurgeon. None of its members acted as readers for the study.

For each case, demographic data, histopathology, and a pain description with an emphasis on neurologic signs were provided to simulate information typically provided to any physician in routine practice. All MR imaging had been performed with a 1.5-T unit (Magnetom Symphony; Siemens, Erlangen, Germany) with a spinal matrix coil. The recruiting radiologist selected 2 images per patient: a sagittal T2-weighted turbo spin-echo sequence (4000/115; section thickness, 4 mm) and an axial T2-weighted turbo spin-echo sequence (4500/110; section thickness, 5 mm) at maximal ESCC grade.[3] The sagittal MRI image included at least 2 spine segments.[13]

The first 90 cases that complied with inclusion criteria were selected. Inclusion criterion was presentation with stage IV (AJCC classification, 7th Edition, 2010) metastatic spine disease confirmed with biopsy. Exclusion criteria included clinical history lacking data required to assess ESCC or imaging of insufficient quality to assess the spinal levels affected.

Procedure

The recruiting radiologist prepared an information pack corresponding to each patient, comprising 2 images and a clinical vignette that included the patient's age, oncologic history, clinical signs, and symptoms (Figure 2).[3] Patient identity was masked and a code was assigned to each information pack. All of the information packs were uploaded to an online platform designed for this study (http://www.typeform.com/).

Each reader was provided with a personal password to assess the information packs online. Readers were asked to indicate all the spinal segments in which they identified metastases for each patient (cervical, thoracic, lumbar, and/or sacral) and to calculate the ESCC score. They were only provided with definitions included in the ESCC (Figure 1). No attempt was made to further define or standardize the meaning of the terms included in the scoring systems or to homogenize the diagnostic criteria, and readers did not receive any instructions regarding the interpretation of images. They were told to use their own clinical judgment when in doubt.

Readers assessed the information pack on their own and uploaded the resulting report directly onto the online platform. They assessed the same clinical sets twice, with a minimum interval period of 6 weeks. The platform software ensured that the minimum interval period was observed, and that readers had no access to their own previous reports or to their colleagues' uploaded reports.

Data introduced into the platform were automatically converted into a spreadsheet. The software engineer in charge of developing the platform cross-checked that the data in the database matched the information introduced into the platform by the readers.

Figure 2. An example of the information pack provided to readers for each patient.

Statistical Analysis

At the analysis phase, grades 1a, 1b, and 1c were grouped, resulting in a 4-point ESCC: 0, 1 (including 1a, 1b, and 1c), 2, and 3. Data on the spine level affected in each patient was classified as cervical, thoracic, lumbar, or sacral, and rated as yes or no.

To assess intraobserver and interobserver agreement, ratings from each reader were cross-tabulated and the Fleiss κ statistic was calculated. A weighted κ approach, with a bisquared weighting scheme, was used to analyze the agreement when using the ESCC scoring system. Kappa values were categorized as "almost perfect" (0.81–1.00), "substantial" (0.61–0.80), "moderate" (0.41–0.60), "fair" (0.21–0.40), "slight" (0.00–0.20), and "poor" (<0.00).[14]

To assess intraobserver agreement for each variable (ESCC score and levels involved), a κ index was calculated for each of the 83 readers, and median, 5th and 95th percentiles values were calculated.

To assess interobserver agreement, the corresponding κ index was calculated and the 95% CI was determined following the jackknife resampling method.[15]

Subgroup analyses for each variable were performed, in which κ values were calculated separately depending on medical specialty, hospital category, and professional experience. Professional experience was classified as "junior" (≤7 years in practice, after residency), "experienced" (8–13 years), and "senior specialist" (≥14 years).

The ESCC scores established by the tumor board were subsequently classified into grades 0, 1, 2, and 3. These grades were used as the gold standard to assess overall agreement. The agreement between this gold standard and the median score for each image among the 83 readers was calculated through the κ statistic.

Stata 13 software was used (StataCorp 2013; Stata Statistical Software: Release 13, College Station, TX).

Results

Of the 132 clinicians invited to act as readers, 83 (62.87%) participated in this study. The first 90 patients selected by the recruiting radiologist (51 women and 39 men; mean age, 60.8 years) complied with the inclusion criteria, and none were excluded. These 90 patients presented metastases in 182 spinal segments. Table 1 shows sample characteristics.

Table 1. Sample Characteristics

Characteristic	n (%)
Hospitals	44
Degree of complexity[a]	
Category 2	3 (6.8)
Category 3	11 (25.0)
Category 4	9 (20.4)
Category 5	21 (47.7)
Management	
Nonprofit[b]	38
For profit[c]	6
Departments	61
Radiology	19 (31.1)
Radiation oncology	11 (18.0)
Orthopedic surgery	12 (19.7)
Neurosurgery	12 (19.7)
Medical oncology	7 (11.5)
Readers	83
Specialty	
Radiology	23 (27.7)
Radiation oncology	22 (26.5)
Orthopedic surgery	16 (19.3)
Neurosurgery	14 (16.9)
Medical oncology	8 (9.6)
Years in practice (post-residency)	
Junior	27 (32.5)
Experienced	25 (30.1)
Senior specialist	31 (37.4)
Setting	
Category of hospital in which they work[a]	
Category 2	3 (3.6)
Category 3	25 (30.1)
Category 4	19 (22.9)
Category 5	36 (43.4)
Hospital management	
Nonprofit[b]	71
For profit[c]	12
Patients	90
Age, y [mean (SD)]	60.8 (12.3)
Male sex	39 (43.3)
Spine segments showing ESCC[d]	
Cervical	12 (13.3)
Thoracic	52 (57.8)
Lumbar	26 (28.9)
ESCC grades[d]	
0	14 (15.6)
1	45 (50.0)
1a	7 (7.8)
1b	20 (22.2)
1c	18 (20.0)
2	21 (23.3)
3	10 (11.1)

Abbreviation: ESCC, epidural spinal cord compression.
[a]Category of hospital; complexity (eg, based on size, availability of high-tech medical equipment and procedures, education activity) ranges from category 1 (the simplest; none of this type were included in this study) to category 5 (the most complex). See text for details.
[b]Hospitals belonging to the Spanish National Health Service (SNHS) or to charities working for the SNHS.
[c]Hospitals under private ownership and management.
[d]According to tumor board and where therapeutic treatment was planned.

There were more than 5 readers for each specialty and degree of professional experience. However, only 3 readers worked at category 2 hospitals; therefore, agreement for this subgroup was not calculated (Tables 2 and 3).

Regarding the identification of spine levels showing ESCC, intraobserver and interobserver agreements were substantial (κ value: median, 0.772 [5th, 95th percentiles: 0.541, 0.948], and κ value: 0.610 [95% CI, 0.531, 0.696], respectively) (Table 2). Subgroup analyses showed that interobserver agreement was only moderate among junior specialists, those working in category 3 hospitals, and in all specialties except radiation oncology (Table 2).

Regarding ESCC score, intraobserver agreement was almost perfect (κ value: median, 0.819 [5th, 95th percentiles: 0.636, 0.923]), and interobserver agreement was substantial (κ value: 0.635 [95% CI, 0.578, 0.699]). Subgroup analyses showed that intraobserver agreement was only substantial among junior specialists, orthopedic surgeons, medical oncologists, and radiation oncologists, whereas interobserver agreement was only moderate among junior and experienced specialists, orthopedic surgeons, and medical oncologists, and among those working in category 3 and 4 hospitals (Table 3).

The agreement between the median of the scores calculated by the readers, and the ESCC grades based on the scores established by the tumor board, was 0.713 (95% CI, 0.596–0.835). Classification by readers and by the tumor board coincided in all of the 31 patients in whom compressive findings were observed (ESCC 2 and 3) (Table 4).

Discussion

Results from this study show that there is a substantial interobserver agreement in determining the ESCC score. These results are generally consistent; differences across specialties, number of years of experience, and type of hospital are small. Results from this study show that there is substantial interobserver agreement in determining the ESCC score. Although some differences in κ values across hospitals, specialties, and number of years of experience were documented, the 95% CI of these values overlap, and differences are small and likely to be clinically meaningless (Tables 2 and 3).[10]

Table 2. Intraobserver and Interobserver Agreement on Level of Cord Compression

	Intraobserver Agreement[a]	Interobserver Agreement[b]
Global agreement	0.772 (0.541; 0.948)	0.610 (0.531; 0.696)
Subgroup analyses		
By specialty		
Orthopedic surgery	0.767 (0.541; 0.882)	0.479 (0.221; 0.781)
Neurosurgery	0.768 (0.589; 0.996)	0.547 (0.319; 0.821)
Medical oncology	0.612 (0.522; 0.806)	0.448 (0.329; 0.667)
Radiation oncology	0.747 (0.508; 0.884)	0.720 (0.597; 0.867)
Radiology	0.841 (0.672; 0.959)	0.576 (0.377; 0.791)
By years of practice		
Junior	0.751 (0.522; 0.903)	0.513 (0.361; 0.689)
Experienced	0.752 (0.541; 0.963)	0.673 (0.552; 0.816)
Senior specialist	0.790 (0.646; 0.943)	0.609 (0.464; 0.770)
By setting (category of hospital)[c]		
Category 2[d]	---	---
Category 3	0.752 (0.522; 0.959)	0.591 (0.458; 0.752)
Category 4	0.722 (0.541; 0.996)	0.667 (0.523; 0.843)
Category 5	0.780 (0.590; 0.943)	0.626 (0.499; 0.768)

[a]κ values: median (5th; 95th percentiles).
[b]κ values (95% CI).
[c]Complexity (eg, based on size, availability of high tech medical equipment and procedures, education activity) ranges from category 1 (the simplest; none of this category were included in this study) to category 5 (the most complex). See text for details.
[d]Only 3 specialists working in category 2 hospitals participated in this study. Therefore, agreement was not calculated for this subgroup.

The substantial agreement in identifying the spinal level showing ESCC based on MRI is reassuring, because clinical symptoms are unreliable for selecting the target level.[13]

This study assessed the reliability of the ESCC score across the different specialties involved in the management of ESCC, in conditions as close as possible to routine practice. All patients showed lesions at 2 or more spine levels and clinicians had to identify the target vertebra based on clinical judgment, as is often the case in clinical practice.[16] A high number of readers participated, and they had differ-

Table 3. Intraobserver and Interobserver Agreement on Spinal Cord Compression Score

	Intraobserver Agreement[a]	Interobserver Agreement[b]
Global agreement	0.819 (0.636; 0.923)	0.635 (0.578; 0.699)
Subgroup analyses		
By specialty		
Orthopedic surgery	0.788 (0.567; 0.972)	0.484 (0.328; 0.692)
Neurosurgery	0.828 (0.723; 0.991)	0.689 (0.571; 0.861)
Medical oncology	0.697 (0.498; 0.840)	0.486 (0.334; 0.726)
Radiation oncology	0.766 (0.639; 0.884)	0.626 (0.533; 0.753)
Radiology	0.859 (0.806; 0.928)	0.682 (0.572; 0.823)
By years of practice		
Junior	0.789 (0.567; 0.885)	0.594 (0.495; 0.720)
Experienced	0.827 (0.615; 0.923)	0.595 (0.501; 0.717)
Senior specialist	0.828 (0.654; 0.969)	0.678 (0.582; 0.799)
By setting (category of hospital)[c]		
Category 2[d]	---	---
Category 3	0.816 (0.567; 0.871)	0.593 (0.493; 0.720)
Category 4	0.817 (0.615; 0.991)	0.564 (0.442; 0.726)
Category 5	0.819 (0.645; 0.923)	0.687 (0.598; 0.798)

[a]κ values: median (5th; 95th percentiles).
[b]κ values (95% CI).
[c]Complexity (eg, based on size, availability of high tech medical equipment and procedures, education activity) ranges from category 1 (the simplest; none of this category were included in this study) to category 5 (the most complex). See text for details.
[d]Only 3 specialists working in category 2 hospitals participated in this study. Therefore, agreement was not calculated for this subgroup.

Table 4. Cross-Tabulation of Scores Determined by ESCC Board Tumor, and Median Categorization of Readers

	Board Tumor				
Score	0	1	2	3	Total
Median ESCC					
0	14	6	0	0	20
1	0	33	1	0	34
2	0	5	18	3	26
3	0	1	2	7	10
Total	14	45	21	10	90

Abbreviation: ESCC, epidural spinal cord compression.

ent backgrounds and worked in different hospitals located in different regions; most readers had never met their colleagues in person. Contrary to some previous studies, the present one did not implement any measures to improve agreement,[17] such as training, consensus, offering a stipend to readers, agreeing on diagnostic criteria, or using standardized nomenclature linked to examples available online.[18] However, results are consistent with those from studies in which spine surgeons used the same scoring system,[3] and those in which radiosurgery experts used an ad hoc version of the Weinstein-Boriani-Biagini scoring system.[19] These results support the use of scoring systems, because the intraobserver and interobserver agreement on the size, location, and shape of tumors is very low when MRI images are analyzed without using such systems, irrespective of physicians' specialty.[20,21]

The similarity of results obtained by physicians, irrespective of years of experience and hospital category, is also generally consistent with previous studies and supports current routine practice[19]; patients with cancer in whom ESCC is suspected undergo MRI at their hospital and are referred to surgery when deemed appropriate.[1] It is reassuring that all of the patients who experienced compressive grades of ESCC (grades 2 and 3), and therefore required urgent clinical management, were correctly identified in this study (Table 4).[19,22] This is consistent with previous studies that showed the sensitivity of the ESCC scoring system for detecting such cases.[3] However, this does not necessarily imply a perfect external validity, because gold standards are difficult to define for metastatic spine disease,[23,24] and discrepancies between imaging and real surgical outcomes do exist.[7]

Good communication among clinicians involved in the management of spine metastatic disease leads to consistency of care, which is a prerequisite for effectiveness.[22] For instance, good communication between surgeons and radiation oncologists facilitates rapid identification of patients with epidural disease in whom surgical resection improves results from postoperative stereotactic body radiotherapy (SBRT).[4] Ensuring that the diagnostic instruments used are reliable, is probably the most effective means of decreasing inappropriate variability in health care.[25] Results from this study suggest that using the ESCC score can be useful to ensure accurate communication among multidisciplinary team members and, therefore, should be used routinely.[6] However, it should be kept in mind that the intrinsic characteristics of certain tumors make it impossible to reach good agreement when it comes to their assessment and management, even after repeated training.[26] Furthermore, agreement when using a scoring system does not necessarily mean that the resulting recommended treatment is appropriate, because clinicians sometimes agree on measures that are not evidence-based or effective,[27] and an improvement in the quality of oncologic care does not necessarily translate immediately into better clinical results or improved survival rates.[28] In fact, no current scoring system is robust enough to establish a solid prognosis for all patients with spinal metastases.[29]

This study has some limitations. Readers were only provided with 2 selected images per case. It is possible that providing all of the readers with all of the images available for each patient might have changed the degree of agreement. However, providing a selection of images ensures that all of the readers assess the same ones, and is consistent with the procedure followed by high-quality studies assessing reliability.[3,30] Readers were volunteers from each of the hospital departments participating in this study, and were not randomly selected. Therefore, selection bias may exist; it is possible that physicians who agreed to participate in this study were the most motivated or interested in metastatic spine disease.[31] However, clinicians involved in management of spinal metastases in routine clinical practice are usually highly specialized, and this does not challenge the results from this study. The prevalence of patients with grades 1a, 1b, and 1c ESCC made it necessary to merge these categories into a single category (grade 1). Maintaining the 3 subcategories would have led to groups too small for the κ statistic to reliably reflect the degree of agreement.[32] In fact, subscale analysis of grade 1 ESCC was not performed in this study or in its original design.[3] Moreover, there are no guidelines on the dosage of irradiation suitable for such cases[19]; for instance, more careful planning of radiotherapy is required for grade 1c ESCC than for grades 1a and 1b in order to avoid reaching the dosage above which the risk of spinal cord radiation overdose and myelopathy increases significantly.[3,33]

Conclusions

This study suggests that there is substantial agreement among radiologists, medical oncologists, radiation oncologists, orthopedic surgeons, and neurosurgeons when identifying the spinal level affected by metastases and when using the ESCC scoring system. Therefore, although there is room for improvement, the use of the ESCC score in clinical practice could improve communication among specialists involved in the management of spine metastases.

Acknowledgments

The authors wish to thank the following investigators of Spanish Back Pain Research Network Task Force for the improvement of interdisciplinary management of spinal metastasis (see supplemental eAppendix 1). The authors are also grateful to Prof. David Moratal for his continuous collaboration.

References

1. Sutcliffe P, Connock M, Shyangdan D, et al. A systematic review of evidence on malignant spinal metastases: natural history and technologies for identifying patients at high risk of vertebral fracture and spinal cord compression. Health Technol Assess 2013;17:1–274.
2. Schmidt MH, Klimo P, Vrionis FD. Metastatic spinal cord compression. J Natl Compr Cancer Netw 2005;3:711–719.
3. Bilsky MH, Laufer I, Fourney DR, et al. Reliability analysis of the epidural spinal cord compression scale. J Neurosurg Spine 2010;13:324–328.
4. Al-Omair A, Masucci L, Masson-Cote L, et al. Surgical resection of epidural disease improves local control following postoperative spine stereotactic body radiotherapy. Neuro Oncol 2013;15:1413–1419.
5. L'espérance S, Vincent F, Gaudreault M, et al. Treatment of metastatic spinal cord compression: CEPO review and clinical recommendations. Curr Oncol 2012;19:e478–490.
6. Arana E, Kovacs FM, Royuela A, et al. Agreement in the assessment of metastatic spine disease using scoring systems. Radiother Oncol 2015;115:135–140.
7. Putz C, Gantz S, Bruckner T, et al. Preoperative scoring and limits of prognostication: functional outcome after surgical decompression in metastatic spinal cord compression. Oncology 2014;86:177–184.
8. Brooks FM, Ghatahora A, Brooks MC, et al. Management of metastatic spinal cord compression: awareness of NICE guidance. Eur J Orthop Surg Traumatol 2014;24(Suppl 1):S255–259.
9. McGivern UM, Drinkwater KJ, Clarke JI, Locke I. A Royal College of Radiologists national audit of radiotherapy in the treatment of metastatic spinal cord compression and implications for the development of acute oncology services. Clin Oncol (R Coll Radiol) 2014;26:453–460.
10. Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011;64:96–106.
11. Donner A, Rotondi MA. Sample size requirements for interval estimation of the kappa statistic for interobserver agreement studies with a binary outcome and multiple raters. Int J Biostat 2010;6:Article 31.
12. [Spanish hospitals classification by cluster analysis.] 2007. Available at: http://www.icmbd.es/docs/resumenClusterHospitales.pdf. Accessed December 7, 2015.
13. Venkitaraman R, Sohaib SA, Barbachano Y, et al. Frequency of screening magnetic resonance imaging to detect occult spinal cord compromise and to prevent neurological deficit in metastatic castration-resistant prostate cancer. Clin Oncol 2010;22:147–152.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159.
15. Efron B. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. Biometrika 1981;68:589–99.
16. Keil S, Barabasch A, Dirrichs T, et al. Target lesion selection: an important factor causing variability of response classification in the Response Evaluation Criteria for Solid Tumors 1.1. Invest Radiol 2014;49:509–517.
17. Obuchowski NA. How many observers are needed in clinical studies of medical imaging? AJR Am J Roentgenol 2004;182:867–869.
18. Fisher CG, Schouten R, Versteeg AL, et al. Reliability of the Spinal Instability Neoplastic Score (SINS) among radiation oncologists: an assessment of instability secondary to spinal metastases. Radiat Oncol 2014;9:69.
19. Cox BW, Spratt DE, Lovelock M, et al. International Spine Radiosurgery Consortium consensus guidelines for target volume definition in spinal stereotactic radiosurgery. Int J Radiat Oncol Biol Phys 2012;83:e597–605.
20. Wu DH, Mayr NA, Karatas Y, et al. Interobserver variation in cervical cancer tumor delineation for image-based radiotherapy planning among and within different specialties. J Appl Clin Med Phys 2005;6:106–110.
21. Dobbs HJ. Defining the radiation target on a daily basis. Cancer Imaging 2006;6:30–32.
22. Terezakis SA, Heron DE, Lavigne RF, et al. What the diagnostic radiologist needs to know about radiation oncology. Radiology 2011;261:30–44.
23. Bankier AA, Levine D, Halpern EF, Kressel HY. Consensus interpretation in imaging research: is there a better way? Radiology 2010;257:14–17.
24. Harel R, Angelov L. Spine metastases: current treatments and future directions. Eur J Cancer 2010; 46: 2696–707.
25. Tuijn S, Janssens F, Robben P, van den Bergh H. Reducing interrater variability and improving health care: a meta-analytical review. J Eval Clin Pract 2012;18:887–895.
26. Rock JB, Washington MK, Adsay NV, et al. Debating deposits: an interobserver variability study of lymph nodes and pericolonic tumor deposits in colonic adenocarcinoma. Arch Pathol Lab Med 2014;138:636–642.
27. Dea N, Fisher CG. Evidence-based medicine in metastatic spine disease. Neurol Res 2014;36:524–529.
28. Keating NL, Landrum MB, Lamont EB, et al. Tumor boards and the quality of cancer care. J Natl Cancer Inst 2013;105:113–121.
29. Tokuhashi Y, Uei H, Oshima M, Ajiro Y. Scoring system for prediction of metastatic spine tumor prognosis. World J Orthop 2014;5:262–271.
30. Fourney DR, Frangou EM, Ryken TC, et al. Spinal instability neoplastic score: an analysis of reliability and validity from the spine oncology study group. J Clin Oncol 2011;29:3072–3077.
31. Lim C, Cheung MC, Franco B, et al. Quality improvement: an assessment of participation and attitudes of medical oncologists. J Oncol Pract 2014;10:e408–414.
32. Lantz CA, Nebenzahl E. Behavior and interpretation of the kappa statistic: Resolution of the two paradoxes. J Clin Epidemiol 1996;49:431–434.
33. Sahgal A, Weinberg V, Ma L, et al. Probabilities of Radiation Myelopathy Specific to Stereotactic Body Radiation Therapy to Guide Safe Practice. Int J Radiat Oncol 2013;85:341–347.

See JNCCN.org for supplemental online content.

JNCCN
Journal of the National Comprehensive Cancer Network

Supplemental online content for:

Agreement in Metastatic Spinal Cord Compression

Estanislao Arana, MD, MHE, PhD; Francisco M. Kovacs, MD, PhD; Ana Royuela, PhD; Beatriz Asenjo, MD, PhD; Úrsula Pérez-Ramírez, MSc; Javier Zamora, PhD; and the Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis

J Natl Compr Canc Netw 2016;14(1):70–76

- **eAppendix 1:** Members of the Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis

eAppendix 1: Members of the Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis (in alphabetical order)

Víctor Abraira[1,2,3]; Lucía Alcázar[1,4]; Ana Alonso[1,5]; Luis Álvarez[1,6]; Marco Antonio Álvarez[1,7]; Guillermo Amengual[1,8]; Aida Antuña[1,9]; Fernando Aparici[1,10]; Andrés Barriga[1,11]; María Barrios[1,12]; Paloma Bas[1,10]; José Begara[1,13]; Francisco Bravo-Rodríguez[1,14]; Alberto Cabrera[1,15]; Carlos Casillas[1,16]; Gregorio Catalán[1,17]; Antonio José Conde[1,16]; Ramón de las Peñas[1,16]; Laura Díaz[1,18]; Diego Dualde[1,19]; Ana Estremera[1,8]; Joaquín Fenollosa[1,12]; Carlos Fernández[1,20]; Eva Fernández[1,21]; Nicomedes Fernández-Baillo[1,22,23]; Pilar Ferrer[1,24]; Salvador Fuster[1,25]; María Isabel Galarraga[1,26]; Cristina García-Villar[1,18]; Luis García-Ferrer[1,27]; Sara García-Duque[1,23]; Javier Garde[1,27]; Andrés González[1,28]; Rafael González-Díaz[1,29]; Alberto Hernández-Fernández[1,30]; Ovidio Hernando[1,23]; Raúl Hernanz[1,21]; Asunción Hervás[1,21]; Esther Holgado[1,23]; María José Juan[1,12]; Javier Lavernia[1,12]; Antonio Lazo[1,31]; Ana Lersundi[1,30]; Escarlata López[1,32]; Margarita Majem[1,33]; Antonio Martín-Benlloch[1,34]; María Isabel Martín[1,12]; Javier Martínez[1,21]; Julia Montoya[1,6]; Paloma Moreno[1,13]; Arturo Navarro[1,35]; Esther Noguerón[1,9]; Ana Ortiz de Mendivil[1,23]; Julio César Palomino[1,36]; Juan Carlos Paniagua[1,37]; David Pereira[1,25], Luis A. Pérez-Romasanta[1,37]; Rocío Pérez[1,38]; Ángel Ramón Piñera[1,6]; Pilar Piñero[1,39]; Julio Plata-Bello[1,40]; José Poblete[1,26]; José Ramírez[1,9]; Daniel Rivas[1,32]; Héctor Roldán[1,40]; Fernando Ruiz[1,41]; José Miguel Sánchez[1,22,23]; Helena Sarasíbar[1,8]; Juan Manuel Sepúlveda[1,42]; Antonio Silvestre[1,19]; Beatriz Sobrino[1,6]; Félix Tomé-Bermejo[1,6]; Isabel Tovar[1,41]; María del Carmen Vallejo[1,21]; Vicente Vanaclocha[1,43]; Asunción Villanueva[1,38]; Joaquín Zamarro[1,44]; and Idoya Zazpe[1,45]

[1]Spanish Back Pain Research Network, Kovacs Foundation, Palma de Mallorca, Spain
[2]CIBER Epidemiology and Public Health (CIBERESP), Madrid, Spain
[3]Clinical Biostatistics Unit, Hospital Ramón y Cajal, IRYCIS, Madrid, Spain
[4]Hospital Universitario la Princesa, Madrid, Spain
[5]Hospital Universitario Rey Juan Carlos, Móstoles, Madrid, Spain
[6]Hospital Universitario Fundación Jiménez Díaz, Madrid, Spain
[7]Hospital Universitario Central de Asturias, Asturias, Spain
[8]Hospital Son Llàtzer, Palma de Mallorca, Spain
[9]Complejo Hospitalario Universitario de Albacete, Albacete, Spain
[10]Hospital Universitario y Politécnico La Fe, Valencia, Spain
[11]Hospital Nacional de Parapléjicos, Toledo, Spain
[12]Fundación Instituto Valenciano de Oncología, Valencia, Spain
[13]Instituto Oncológico Xanit, Benalmádena, Málaga, Spain
[14]Hospital Universitario Reina Sofía, Córdoba, Spain
[15]OSATEK. Hospital de Galdakao, Vizcaya, Spain
[16]Consorcio hospitalario Provincial de Castellón, Castellón, Spain
[17]Hospital Universitario Cruces, Barakaldo, Vizcaya, Spain
[18]Hospital Universitario Puerta del Mar, Cádiz, Spain
[19]Hospital Clínico Universitario de Valencia, Valencia, Spain
[20]Hospital General Universitario Gregorio Marañón, Madrid, Spain
[21]Hospital Universitario Ramón y Cajal, Madrid, Spain
[22]Hospital Universitario la Paz, Madrid, Spain
[23]HM Universitario Sanchinarro, Madrid, Spain
[24]Hospital Intermutual de Levante, San Antonio de Benagéber, Valencia, Spain
[25]Hospital Clinic de Barcelona, Barcelona, Spain
[26]Hospital de Manacor, Manacor, Islas Baleares, Spain
[27]Hospital Arnau de Vilanova, Valencia, Spain
[28]Hospital Universitario Marqués de Valdecilla, Santander, Spain
[29]Hospital Universitario Fundación Alcorcón, Alcorcón, Madrid, Spain
[30]Hospital Universitario Donostia, Donostia, Spain
[31]Clínica Oncosur, Córdoba, Spain
[32]Clínica Oncosur, Granada, Spain
[33]Hospital de la Santa Creu i Sant Pau, Barcelona, Spain
[34]Hospital Doctor Peset, Valencia, Spain
[35]Hospital Duran i Reynals, l'Hospitalet de Llobregat, Barcelona, Spain
[36]Hospital POVISA, Vigo, Spain
[37]Hospital Universitario de Salamanca, Salamanca, Spain
[38]Hospital Regional Universitario Carlos Haya, Málaga, Spain
[39]Hospital Universitario Virgen del Rocío, Sevilla, Spain
[40]Hospital Universitario de Canarias, Islas Canarias, Spain
[41]Hospital Universitario Virgen de las Nieves, Granada, Spain
[42]Hospital Universitario 12 de Octubre, Madrid, Spain
[43]Hospital 9 de Octubre, Valencia, Spain
[44]Hospital Universitario Virgen de la Arrixaca, Murcia, Spain
[45]Complejo Hospitalario de Navarra, Navarra, Spain

3. Escala de inestabilidad neoplásica de columna vertebral . Acuerdo entre distintas especialidades médicas y quirúrgicas

The Spine Journal 2016;16:591–599

Antecedentes y contexto: la inestabilidad de columna vertebral es una complicación conocida de las metástasis; a pesar de los criterios recientemente sugeridos, no está claramente definida en la bibliografía.

Propósito: Este estudio se dirigió a comprobar el acuerdo intra- e interobservdaor cuando se usa la escala de inestabilidad neoplásica de columna vertebral (SINS, en sus acrónimo inglés) por todos los médicos involucrados en su tratamiento.

Diseño de estudio: Estudio de fiabilidad multicéntrico independiente para el recientemente creado SINS, realizado con un panel de oncología médica, neurocirugía, radiología, cirugía ortopédica y oncología radioterápica.

Medidas de resultados: Se empleó el coeficiente de correlación intraclase (CCI) para el acuerdo de la escala SINS. El kappa de Fleiss se usó para comprobar el acuerdo en la localización del nivel vertebral más afectado, el acuerdo en la categoría SINS (estable, potencialmente estable o inestable); y el acuerdo global con la clasificación establecida por el comité de tumores.

Métodos: Los datos de imagen y clínicos de 90 pacientes con metástasis demostradas por biopsia se les proporcionaron a 83 especialista de 44 hospitales. No se establecieron criterios prestablecidos. Cada médicos determinó la escala SINS dos veces, con un intervalo mínimo de 6 semanas. Los médicos estaban ciegos a cada evaluación suya y a la de los otros médicos. Los análisis de subgrupos se realizaron según la especialidad clínica (oncología médica, neurocirugía, radiología, cirugía ortopédica y oncología radioterápica), años de experiencia (≤ 7, 8-13, ≥ 14), y tipo de hospital (cuatro niveles según tamaño y complejidad). Este estudio fue apoyado por la Fundación Kovacs.

Resultados: el acuerdo intra- e interobservador en la localización de los niveles más afectados fue casi perfecto (k>0,94). El acuerdo intra-observador en la escala SINS fue excelente (CCI=0,77), mientras el acuerdo interobservador fue moderado (CCI=0,55). El acuerdo intra-observador en la categoría SINS fue sustancial (k=0,61), mientras el acuerdo interobservador fue moderado (k=0,42). El acuerdo global con la clasificación

del comité de tumores fue sustancial (k=0,61). Los resultados fueron similares entre especialidades, años de experiencia y categoría hospitalaria.

Conclusiones:

El acuerdo en el estudio de la inestabilidad metastásica de la columna vertebral es moderado. La escala SINS puede mejorar la comunicación entre los médicos de asistencia oncológica.

The Spine Journal 16 (2016) 591–599

THE SPINE JOURNAL

Clinical Study

Spine Instability Neoplastic Score: agreement across different medical and surgical specialties

Estanislao Arana, MD, MHE, PhD[a,b,c,*], Francisco M. Kovacs, MD, PhD[c,d], Ana Royuela, PhD[c,e,f], Beatriz Asenjo, MD, PhD[c,g], Úrsula Pérez-Ramírez, MSc[c,h], Javier Zamora, PhD[c,e,f,i] the Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis

[a]*Department of Radiology, Valencian Oncology Institute Foundation, C/ Beltrán Báguena, 19, 46009 Valencia, Spain*

[b]*Research Institute in Health Services Foundation, C/ San Vicente,112, 3, 46007 Valencia, Spain*

[c]*Spanish Back Pain Research Network, Kovacs Foundation, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain*

[d]*Scientific Department, Kovacs Foundation, Paseo de Mallorca 36, 07012 Palma de Mallorca, Spain*

[e]*CIBER Epidemiology and Public Health (CIBERESP), Av. Monforte de Lemos, 3-5. Pabellón 11. Planta 0, 28029 Madrid, Spain*

[f]*Clinical Biostatistics Unit, Hospital Ramón y Cajal, IRYCIS. Ctra. Colmenar Km. 9.1, 28034 Madrid, Spain*

[g]*Department of Radiology, Hospital Regional Universitario Carlos Haya, Avda Carlos Haya s/n, 29010 Málaga, Spain*

[h]*Center for Biomaterials and Tissue Engineering, Universitat Politècnica de València, CPI Building (8E), F access, 1st floor, Cami de Vera, s/n, 46022 Valencia, Spain*

[i]*Barts and the London School of Medicine & Dentistry, Queen Mary University of London, Mile End Road, London E1 4NS, UK*

Received 25 March 2015; revised 27 August 2015; accepted 6 October 2015

Abstract

BACKGROUND CONTEXT: Spinal instability is an acknowledged complication of spinal metastases; in spite of recent suggested criteria, it is not clearly defined in the literature.

PURPOSE: This study aimed to assess intra and interobserver agreement when using the Spine Instability Neoplastic Score (SINS) by all physicians involved in its management.

STUDY DESIGN: Independent multicenter reliability study for the recently created SINS, undertaken with a panel of medical oncologists, neurosurgeons, radiologists, orthopedic surgeons, and radiation oncologists, was carried out.

PATIENT SAMPLE: Ninety patients with biopsy-proven spinal metastases and magnetic resonance imaging, reviewed at the multidisciplinary tumor board of our institution, were included.

OUTCOME MEASURES: Intraclass correlation coefficient (ICC) was used for SINS score agreement. Fleiss kappa statistic was used to assess agreement on the location of the most affected vertebral level; agreement on the SINS category ("stable," "potentially stable," or "unstable"); and overall agreement with the classification established by tumor board.

METHODS: Clinical data and imaging were provided to 83 specialists in 44 hospitals across 14 Spanish regions. No assessment criteria were pre-established. Each clinician assessed the SINS score twice, with a minimum 6-week interval. Clinicians were blinded to assessments made by other specialists and to their own previous assessment. Subgroup analyses were performed by clinicians' specialty, experience (≤7, 8–13, ≥14 years), and hospital category (four levels according to size and complexity). This study was supported by Kovacs Foundation.

RESULTS: Intra and interobserver agreement on the location of the most affected levels was "almost perfect" (κ>0.94). Intra-observer agreement on the SINS score was "excellent" (ICC=0.77), whereas interobserver agreement was "moderate" (ICC=0.55). Intra-observer agreement in SINS category was "substantial" (k=0.61), whereas interobserver agreement was "moderate" (k=0.42). Overall agreement with the tumor board classification was "substantial" (κ=0.61). Results were similar across specialties, years of experience, and hospital category.

FDA device/drug status: Not applicable.

Author disclosures: ***EA:*** Nothing to disclose. ***FMK:*** Nothing to disclose. ***AR:*** Nothing to disclose. ***BA:*** Nothing to disclose. ***UPR:*** Nothing to disclose. ***JZ:*** Nothing to disclose.

The authors declare that they have no conflicts of interest.

* Corresponding author. Fundación Instituto Valenciano de Oncología, C/ Beltrán Báguena, 19, 46009 Valencia, Spain. Tel.: +34-961114360.

E-mail address: aranae@uv.es (E. Arana)

http://dx.doi.org/10.1016/j.spinee.2015.10.006

CONCLUSIONS: Agreement on the assessment of metastatic spine instability is moderate. The SINS can help improve communication among clinicians in oncology care.

Keywords: Medical specialty; Observer agreement; Reliability analysis; Spinal instability; Spine Instability Neoplastic Score; Spinal metastases

Introduction

The organ most commonly affected by metastatic cancer is the skeleton, which is also where it causes the highest morbidity [1]. There is controversy on the exact definition of spinal instability caused by spine metastatic disease, and its appropriate management [2]. Several scoring systems have been proposed to standardize the diagnosis of "spinal instability" in these patients, and selecting those in whom surgery should be considered [3–5]. However, only 14% of British clinicians managing spine metastatic disease are familiar with the available scoring systems [6].

The Spine Instability Neoplastic Score (SINS) is based on clinical data and imaging findings (Table 1), and has been suggested as the most straightforward scoring system [7]. It was originally developed by spine surgeons, and very few studies have analyzed its reliability when used by different specialists [8–10]. Higher SINS score has been shown as predictor of radiotherapy failure [11]. None have included physicians from all the specialties involved in the management of spine metastatic disease.

Table 1
The SINS classification according to the Spine Oncology Study Group (SOSG) [7]

Location	Score
Junctional (occiput–C2, C7–T2, T11–L1, L5–S1)	3
Mobile spine (C3–C6, L2–L4)	2
Semirigid (T3–T10)	1
Rigid (S2–S5)	0
Pain*	
Yes	3
Occasional pain but not mechanical	1
Pain-free lesion	0
Bone lesion	
Lytic	2
Mixed (lytic/blastic)	1
Blastic	0
Radiographic spinal alignment	
Subluxation/translation present	4
De novo deformity (kyphosis/scoliosis)	2
Normal alignment	0
Vertebral body collapse	
>50% collapse	3
<50% collapse	2
No collapse with>50% body involved	1
None of the above	0
Posterolateral involvement of spinal elements†	
Bilateral	3
Unilateral	1
None of the above	0

* Pain improvement with recumbency or pain with movement or loading of spine.
† Facet, pedicle, or costovertebral joint fracture or replacement with tumor.

Assessing the reliability of SINS across the different specialists involved in the assessment of spine metastatic disease may contribute to improving the decision-making process on the most suitable treatment for each patient.

Therefore, the purpose of this study was to assess intra- and interobserver agreement in (a) the calculation of the SINS score, (b) the classification of spine instability based on this score, and (c) the location of the most affected vertebral level, in conditions as close as possible to routine clinical practice, among a large sample of clinicians from different specialties with varied degrees of experience and working in different settings and locations.

Methods

Study design and participants

This prospective study was approved by the institutional review boards of the participating hospitals, and complied with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [12].

Selection of hospital departments and clinicians

At the design phase of this study, the medical specialties considered to be relevant for the management of spine metastatic spine disease were listed as follows: neurosurgery, medical oncology, radiation oncology, radiology, and orthopedic surgery.

All of the 61 hospital departments specializing in these clinical areas, which had previously participated in studies undertaken by the American Joint Commitee on Cancer or had expressed interest in doing so, were invited to participate in this study. Twelve departments were located in six private hospitals and the other 49 in 38 non-profit hospitals, belonging to, or working for, the Spanish National Health Service (SNHS). The SNHS is the tax-funded, government-run, organization which provides free health care to every resident in Spain.

The SNHS classifies Hospitals in five categories, based on the size of the catchment area, number of beds, number of clinicians, availability of high tech medical equipment and procedures, education, training and academic activity, and clinical complexity of the cases treated (ie, being the "reference hospital" for specific diseases or procedures) [13]. Category 1 is the simplest and category 5 is the most complex. Departments invited to participate were located in hospitals belonging to categories 2, 3, 4, and 5.

EVIDENCE & METHODS

Context
A number of scoring systems and classifications schemes have been developed to inform the care of patients afflicted with spinal tumors. Many of these have not been independently validated outside of the cohorts used to develop the scoring systems, nor have their inter-rater or intra-rater reliabilities been assessed.

Contribution
The authors assessed the inter- and intra-rater agreement using the Spine Instability Neoplastic Score (SINS) among a heterogeneous group of physicians involved in the management of patients with spinal metastases. This study was conducted within the tumor board of an institution in Spain. The authors conclude that results of their work show that metrics for the SINS category ranged from moderate to substantial agreement between and within raters, respectively. The authors maintain that these findings attest to the utility of the SINS schema in a clinical setting.

Implications
The authors' analysis provides useful information regarding the clinical utility and reliability of the SINS. It should be recognized that among the tumor board at which the study was performed, reviewer familiarity and thought processes developed at the institutional level over time may improve the inter-rater reliability as compared to practitioners utilizing the scheme at different centers. This is a possible line for further inquiry and likely is necessary before a definite characterization of the inter- and intra-rater reliability of the SINS can be accepted.

—*The Editors*

All clinicians who had finished their residency and worked at the participating departments were invited to act as readers in this study. Those who accepted were asked to provide the number of years they had been working in clinical practice after their residency. The departments and clinicians did not receive any compensation for participating in this study.

Selection of patients and images

Patients and images were selected by a radiologist who worked in a category 4 hospital and did not act as a reader in this study. He revised consecutive patients in whom a tumor board (composed by a medical oncologist, a radiation oncologist, an orthopedic surgeon, a radiologist, and a pathologist, none of whom acted as readers in this study) had established the diagnosis of spine metastatic disease at ≥2 spine levels and had assessed the SINS score. These cases were reviewed in reverse chronological order (ie, more recent cases were revised first).

All images were acquired on the same computed tomography (CT) and magnetic resonance imaging (MRI) systems with the same technique. The radiologist selected four images per patient; two CT scans and two MRI images, comprising at least two spine levels.

The first 90 cases that complied with inclusion criterion and not with exclusion criteria were selected. Inclusion criterion was presenting a stage IV (American Joint Commitee on Cancer classification 7th Edition, 2010) biopsy-proven spine metastatic disease. Exclusion criteria were (a) clinical history lacking data required to assess SINS or (b) imaging of insufficient quality to assess the spinal level or levels affected.

Procedure

The recruiting radiologist prepared an information pack on each patient, comprising the four images and a clinical vignette stating patient's age, oncologic history, clinical signs and symptoms, and whether the patient suffered from movement-related pain (Figure 1) [8]. Patient identity was masked and a code was assigned to each information pack. All the information packs were uploaded onto an online platform specifically designed for this study (http://www.typeform.com/).

Each reader was provided with a personal password to access the information packs online. For each patient, readers were asked to report all the spinal levels in which they detected metastases (cervical, thoracic, lumbar, or sacral) and to calculate the SINS score based on the segment which they considered to be most affected (ie, the "target" vertebral level; eg, L1–L2). Readers were only provided with definitions included in the SINS (Table 1). No attempt was made to further explain or standardize these definitions, and readers did not receive any instructions regarding the interpretation of images. They were told to use their own clinical judgment when in doubt, as they would do in everyday, routine clinical practice.

Readers assessed the information pack alone and on their own, and introduced the resulting report into the online platform. They were asked to assess the same clinical sets twice, with a minimum 6-week interval. The software ensured that the minimum period was observed, and that readers had no access to their own previous reports or to their colleagues' uploaded reports.

Data introduced into the platform were automatically transferred into a spreadsheet. The software engineer in charge of developing the platform cross-checked the spreadsheet against the data introduced into the platform by the readers before sending the information to the biostatisticians in charge of statistical analysis.

Statistical analysis

Sample size was calculated at 90 patients with spine metastatic disease, assuming an intraclass correlation coefficient (ICC) of 0.7, a width of the confidence interval (CI) of 0.15, and that at least 5 observers per specialty would be recruited.

To assess agreement in the SINS score, the ICC was calculated using a two-way random-effects model. For intra-observer agreement, an ICC was calculated for each one of the 83 observers, and median and 5th and 95th percentiles were estimated. For interobserver agreement, scores from the first round were analyzed, and the ICC and its 95% CI were estimated. Intraclass correlation coefficient values were categorized as showing reliability to be "excellent" (>0.75), "moderate" (0.4–0.75), or "poor" (<0.4) [14].

The SINS scores were then collapsed into three categories according to the degree of stability they represent and the treatment they imply: "stable" (SINS score between 0 and 6), "potentially unstable" (7–12), or "unstable" (13–18) [7].

The unstable spine levels in each patient were classified into four categories: cervical, thoracic, lumbar, or sacral.

To assess intra-observer agreement for each categorical variable, a Fleiss kappa index was calculated for each one of the 83 readers, and median, 5th, and 95th percentile values were calculated [15]. To assess interobserver agreement, the corresponding kappa index was estimated, and the 95% CI was determined following the jackknife resampling method [16]. A weighted-kappa approach, with a bi-squared weighting scheme, was used. Kappa values were categorized as "almost perfect" (0.81–1.00), "substantial" (0.61–0.80), "moderate" (0.41–0.60), "fair" (0.21–0.40), "slight" (0.00–0.20), and "poor" (<0.00) [17].

Subgroup analyses for each variable were performed, in which ICC and kappa values were calculated separately depending on medical specialty, hospital category, and professional experience. Degree of professional experience was classified as "recently specialized" (≤7 years in practice, after residency), "experienced" (8–13 years), and "senior specialist" (≥14 years).

The SINS scores agreed by the tumor board, and subsequently classified as "stable," "potentially unstable," or "unstable," were used as the "gold standard" to assess overall agreement. The agreement between this gold standard and the median score for each image among the 83 readers was calculated through the kappa statistic.

Statistical package Stata v 13 (StataCorp. 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP) was used.

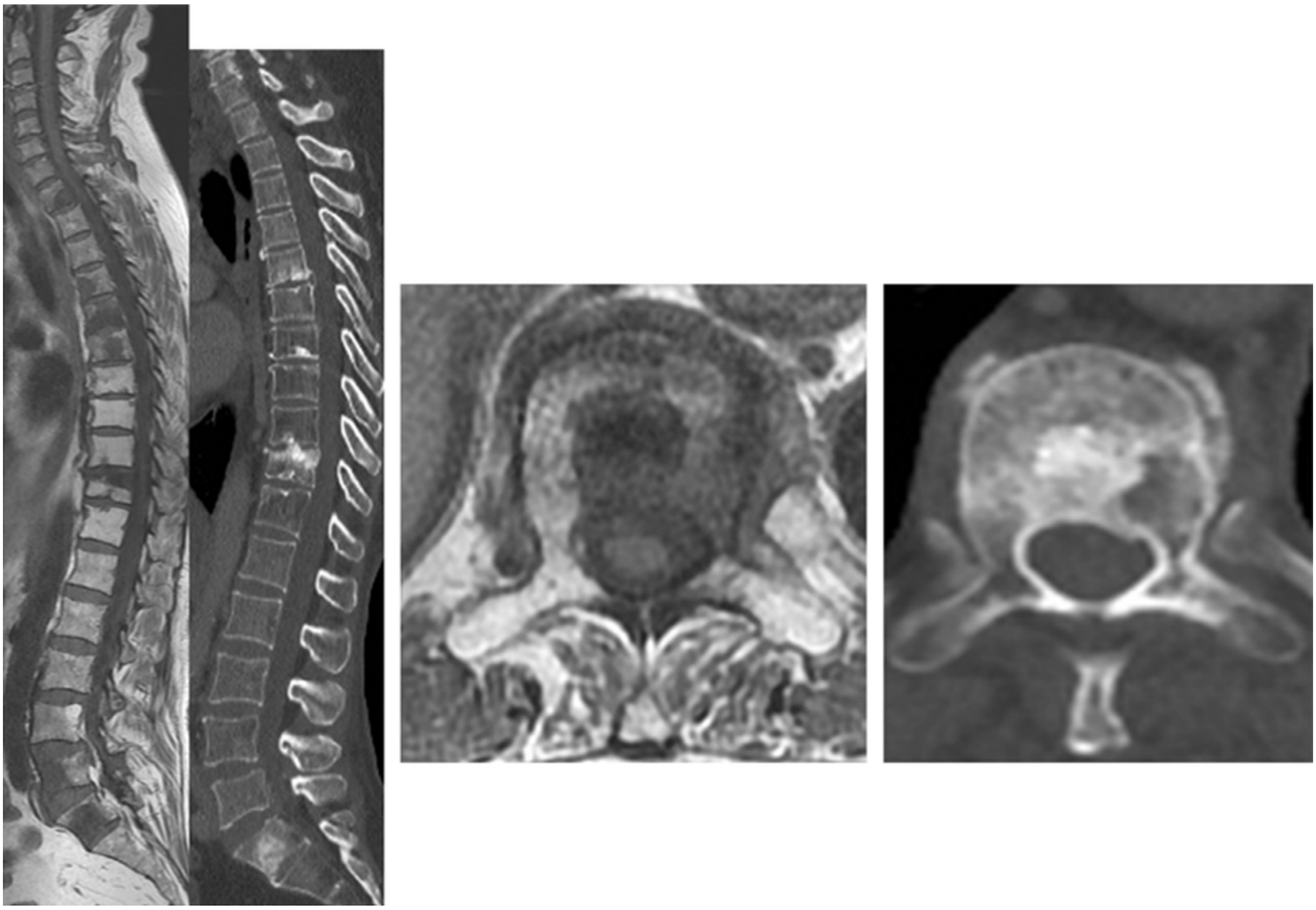


Figure 1. An example of the information pack provided to readers for each patient.
Images corresponding to a 69-year-old woman suffering from breast cancer, who reported continuous back pain without referred pain. She presented lung, liver, and bone metastases. Please select the most unstable spine level and fill in the corresponding SINS scoring.
☐Cervical ☐Thoracic ☐Lumbar ☐ Sacrum
SINS _____

Results

Eighty-three of 132 (62.87%) clinicians who were invited to act as readers participated in this study, and 49 specialists declined (Table 2). The first 90 patients selected by the recruiting radiologist complied with the inclusion criteria, and none were excluded. These 90 patients showed metastases in 182 spinal levels, which originated from 16 primary cancers, with breast (n=37), prostate (n=16), and lung (n=12) being the most common (Table 2).

There were more than five readers for each specialty and degree of professional experience. However, only three readers worked at category 2 hospitals; therefore, agreement for this subgroup was not calculated (Tables 3–5).

Intra-observer agreement on the SINS score was "excellent" (median ICC 0.767; 5th, 95th percentiles [0.538; 0.939]). Interobserver agreement was "moderate" (0.546; 95% CI [0.476; 0.624]). The only exception found in subgroup analyses was that intra-observer agreement was only "moderate" among medical and radiation oncologists, as well as among physicians with 8–13 years of experience (Table 3).

When the SINS scores were grouped into categories ("stable," "potentially unstable," or "unstable"), intra-observer agreement in classifying the patients into these categories was "substantial" (median kappa 0.605; 5th, 95th percentiles [0.381; 0.880]) whereas interobserver agreement was "moderate" (0.424; 95% CI [0.336; 0.524]). Subgroup analyses revealed the following exceptions: (a) intra-observer agreement was only "moderate" among medical oncologists, radiation oncologists, physicians with ≤7 years of experience, and physicians working in hospitals in categories 3 and 5; (b) interobserver agreement was only "fair" among orthopedic surgeons, radiologists, physicians with ≥14 years of clinical experience, and physicians working in category 5 hospitals (Table 4).

Intra- and interobserver agreement in the identification of the potentially unstable spinal level(s), based on the categories grouping the SINS scores, was "almost perfect" (median kappa 0.971; 5th, 95th percentiles [0.871; 1.000] and 0.944; 95% CI [0.922; 0.970], respectively). Subgroup analyses did not show any differences (Table 5).

Overall agreement with the tumor board classification was "substantial" (kappa [95% CI]; 0.610 [0.437; 0.792]) All patients classified by the tumor board as "unstable" were rated with ≥7 SINS points. However, among the 14 patients who were classified as "stable" by the tumor board, nine were rated with a median SINS score suggesting "potentially unstable" (Table 6).

Discussion

Results from this study show that there is a "moderate" interobserver agreement in determining the SINS score and in using this score to classify patients into three categories according to spine stability. They also show that this classification largely matches the consensus-based classification

Table 2
Sample characteristics

Hospitals*		44
	Degree of complexity†	
	Category 2	3 (6.8)
	Category 3	11 (25)
	Category 4	9 (20.4)
	Category 5	21 (47.7)
	Management‡	
	Not for profit	38
	For profit	6
Departments*		61
	Radiology	19 (31.1)
	Radiation oncology	11 (18.0)
	Orthopedic surgery	12 (19.7)
	Neurosurgery	12 (19.7)
	Medical oncology	7 (11.5)
Readers*		83 [49]
	Specialty	
	Radiology	23 (27.7) [14]
	Radiation oncology	22 (26.5) [14]
	Orthopedic surgery	16 (19.3) [10]
	Neurosurgery	14 (16.9) [6]
	Medical oncology	8 (9.6) [5]
	Years in practice (post-residency)	
	≤7	27 (32.5) [14]
	8–13	25 (30.1) [17]
	≥14	31 (37.4) [18]
	Setting	
	Category of hospital in which they work†	
	Category 2	3 (3.6) [1]
	Category 3	25 (30.1) [18]
	Category 4	19 (22.9) [12]
	Category 5	36 (43.4) [18]
	Hospital management‡	
	Not for profit	71 [40]
	For profit	12 [9]
Patients		90
	Age (years)§	60.8 (12.3)
	Gender (males)*	39 (43.3)
	Location of metastases*	
	Cervical	4 (4.4)
	Cervical and thoracic	15 (16.7)
	Cervical, thoracic, and lumbar	1 (1.1)
	Cervical, thoracic, lumbar, and sacral	2 (2.2)
	Thoracic	18 (20)
	Thoracic and lumbar	15 (16.7)
	Thoracic, lumbar, and sacral	24 (26.7)
	Lumbar	5 (5.6)
	Lumbar and sacral	6 (6.7)
Spinal levels analyzed for stability*,‖		
	Cervical	8 (8.9)
	Thoracic	53 (58.9)
	Lumbar	29 (32.2)

* n (%). The number in square brackets indicate number of invited specialists who declined to participate.
† Category of hospital; complexity (based on size, availability of high-tech medical equipment and procedures, education activity, etc.) ranges from category 1 (the simplest—none of this type was included in this study) to category 5 (the most complex). See text for details.
‡ Not for profit: Hospitals belonging to the Spanish National Health Service (SNHS) or to charities working for the SNHS. For profit: Hospitals privately own and managed.
§ Mean (SD).
‖ Assessed by a multi-disciplinary tumor board (see text for details).

Table 3
Intra- and interobserver agreement on SINS score (0–18), as measured by ICC

	Intra-observer agreement*	Interobserver agreement†
Global agreement	0.767 (0.538; 0.939)	0.546 (0.476; 0.624)
Subgroup analyses		
By specialty		
Orthopedic surgery	0.796 (0.456; 0.972)	0.629 (0.557; 0.704)
Neurosurgery	0.763 (0.538; 0.827)	0.566 (0.488; 0.648)
Medical oncology	0.687 (0.000; 0.768)	0.450 (0.364; 0.544)
Radiation oncology	0.724 (0.531; 0.957)	0.513 (0.433; 0.599)
Radiology	0.816 (0.627; 0.889)	0.622 (0.547; 0.699)
By years of practice		
≤7	0.757 (0.456; 0.954)	0.511 (0.437; 0.594)
8–13	0.732 (0.608; 0.880)	0.557 (0.480; 0.639)
≥14	0.799 (0.531; 0.972)	0.565 (0.491; 0.645)
By setting (category of hospital)‡		
Category 2§	—	—
Category 3	0.748 (0.456; 0.854)	0.514 (0.439; 0.597)
Category 4	0.805 (0.538; 0.972)	0.563 (0.485; 0.646)
Category 5	0.760 (0.590; 0.957)	0.556 (0.483; 0.636)

ICC, intraclass correlation coefficient.
* ICC values: median (5th; 95th percentiles).
† Individual ICC value (95% confidence interval).
‡ Complexity (based on size, availability of high-tech medical equipment and procedures, education activity, etc.) ranges from category 1 (the simplest—none of this category was included in this study) to category 5 (the most complex). See text for details.
§ Only three specialists working in category 2 hospitals participated in this study. Therefore, agreement was not calculated for this subgroup.

Table 4
Intra- and interobserver agreement on SINS category among the 83 clinicians, as measured by kappa values

	Intraobserver agreement*	Interobserver agreement†
Global agreement	0.605 (0.381; 0.880)	0.424 (0.336; 0.524)
Subgroup analyses		
By specialty		
Orthopedic surgery	0.675 (0.455; 1.000)	0.399 (0.053; 0.870)
Neurosurgery	0.634 (0.389; 0.825)	0.497 (0.307; 0.753)
Medical oncology	0.509 (0.066; 0.596)	0.429 (0.183; 0.813)
Radiation oncology	0.578 (0.381; 0.937)	0.462 (0.234; 0.759)
Radiology	0.646 (0.460; 0.799)	0.328 (0.205; 0.486)
By years of practice		
≤7	0.594 (0.358; 0.934)	0.410 (0.228; 0.641)
8–13	0.619 (0.423; 0.800)	0.511 (0.329; 0.743)
≥14	0.633 (0.365; 1.000)	0.345 (0.239; 0.477)
By setting (category of hospital)‡		
Category 2§	—	—
Category 3	0.580 (0.353; 0.780)	0.425 (0.245; 0.655)
Category 4	0.665 (0.389; 1.000)	0.530 (0.310; 0.819)
Category 5	0.595 (0.418; 0.937)	0.372 (0.249; 0.523)

* κ values: median (5th; 95th percentiles).
† κ value (95% confidence interval).
‡ Complexity (based on size, availability of high-tech medical equipment and procedures, education activity, etc.) ranges from category 1 (the simplest—none of this category was included in this study) to category 5 (the most complex). See text for details.
§ Only three specialists working in category 2 hospitals participated in this study. Therefore, agreement was not calculated for this subgroup.

Table 5
Agreement in the spinal levels involved, as measured by the kappa statistic

	Intraobserver agreement*	Interobserver agreement†
Global agreement	0.971 (0.871; 1.000)	0.944 (0.922; 0.970)
Subgroup analyses		
By specialty		
Orthopedic surgery	0.956 (0.813; 1.000)	0.923 (0.871; 0.997)
Neurosurgery	0.972 (0.927; 1.000)	0.907 (0.814; 1.000)
Medical oncology	0.909 (0.813; 0.956)	0.894 (0.763; 1.000)
Radiation oncology	0.970 (0.891; 1.000)	0.974 (0.953; 1.000)
Radiology	0.986 (0.944; 1.000)	0.964 (0.930; 1.000)
By years of practice		
≤7	0.971 (0.826; 1.000)	0.908 (0.856; 0.976)
8–13	0.971 (0.926; 1.000)	0.973 (0.953; 0.997)
≥14	0.970 (0.906; 1.000)	0.954 (0.920; 0.999)
By setting (category of hospital)‡		
Category 2§	—	—
Category 3	0.971 (0.871; 1.000)	0.931 (0.892; 0.981)
Category 4	0.972 (0.813; 1.000)	0.973 (0.948; 1.000)
Category 5	0.970 (0.863; 1.000)	0.954 (0.924; 0.994)

* κ values: median (5th; 95th percentiles).
† κ value (95% confidence interval).
‡ Complexity (based on size, availability of high-tech medical equipment and procedures, education activity, etc.) ranges from category 1 (the simplest—none of this category was included in this study) to category 5 (the most complex). See text for details.
§ Only three specialists working in category 2 hospitals participated in this study. Therefore, agreement was not calculated for this subgroup.

established by a multidisciplinary tumor board, and that there is an "almost perfect" agreement in the identification of the unstable spine levels in each patient (Tables 3–6). These results are generally consistent across all the specialties involved in managing spine metastatic disease, irrespective of the number of years of experience and the size and complexity of the hospitals where the specialists work. The excellent agreement in the selection of the target level is reassuring, because disagreement is the major source of variability when assessing oncology patients' individual response to treatment [18].

Some previous studies found the interobserver agreement in the SINS score to be "excellent" [8,10,19–21], whereas the present study only found "moderate" agreement. Differences in methods can account for this; the current study aimed to assess intra- and interobserver agreement in conditions as close as possible to routine clinical practice; all patients showed metastases in at least two spine levels, and identification of the target vertebral level was based on clinical judgment, as in routine practice [18]. Moreover, a high number of readers participated; they had different backgrounds and worked in hospitals that were located in different regions, most readers had never met their colleagues in person, and agreement was assessed among different readers, and not among their individual scores and their global mean score [8,20]. Furthermore, contrary to some previous studies, the present study did not implement any measures to improve agreement [22], such as training, offering a stipend to readers, agreeing on diagnostic criteria, or using standardized nomenclature linked to examples available online [21,23,24].

Table 6
Cross-tabulation of scores determined by SINS Board tumor and median categorization of readers*

		Board tumor			
		Stable (≤6)	Potentially unstable (7–12)	Unstable (≥13)	Total
Median SINS score	Stable (≤6)	5 (35.7 %)	0 (0.0 %)	0 (0.0 %)	5
	Potentially unstable (7–12)	9 (64.3%)	59 (98.3 %)	5 (31.2 %)	73
	Unstable (≥13)	0 (0.0%)	1 (1.7 %)	11(68.8 %)	12
Total		14	60	16	90

* Predictive validity (kappa value): 0.610 (95% CI, 0.437; 0.792).

As opposed to what has been found in this study (Tables 3–5), a previous report found agreement to be higher among physicians with more years of experience [20]. The fact that all physicians who participated in the current study had undergone ≥4 years of clinical training to become certified specialists may account for this difference. Paradoxically, in the current study, the physicians with the highest degree of experience showed the smallest interobserver agreement when their SINS ratings were collapsed into three categories. However, although their median kappa value was smaller than the one for physicians with less experience, the 5th, 95th percentiles largely overlap (Table 4).

The assessment of imaging by spine surgeons is usually considered as the gold standard for deciding whether surgery is appropriate for a patient with metastatic spine disease [19], and a previous study found that the interobserver agreement in the SINS score is higher among spine surgeons than among other specialists [20]. This was not the case in the current study, where differences across specialties were inconsistent, small, and likely to be clinically meaningless (Tables 3–5) [12]. The large sample size in this study, the high number of participating clinicians from each specialty, and the fact that, as opposed to other studies [8,20], none of the readers participated in the definition of the "gold standard," and those who were not spine surgeons were specialists who also manage spine metastatic disease in routine practice, can account for these differences in results.

"Interobserver agreement" does not necessarily mean "external validity," because consensus may not represent the actual "truth" [25]; sometimes clinicians agree on measures which are not evidence-based or effective [26]. In fact, the correlation between imaging and histopathology findings is low in some types of cancer [27], differences between SINS classification and real surgical outcomes have been documented [10], and the intrinsic characteristics of some types of tumor make it impossible to achieve high levels of agreement in clinical decisions [28]. Moreover, "agreement" when using a scoring system does not necessarily mean that the recommended treatment is "appropriate" or that it will improve outcomes.

The degree of agreement among different specialists when using the SINS score, the substantial agreement with the tumor board classification, and the excellent agreement in the selection of the target level, suggest that generalizing the use of the SINS score in routine practice would facilitate good communication among the different specialists involved in the management of spinal metastases. Even though improvement in the quality of care does not necessarily translate immediately into better clinical results [29], good communication among the different specialists involved in the management of oncology patients leads to consistency of care, which is a prerequisite for effectiveness in oncology patients [30].

Future studies should compare the reliability and prognostic validity of different scoring systems, such as the SINS [31] and the Taneichi scores [3,32], and assess whether their use, or measures to improve interobserver agreement, actually lead to improved outcomes.

This study has some potential limitations. Readers only analyzed four selected images per case. Providing all the readers with all the images available for each patient might have changed the degree of agreement. However, this is the usual procedure for assessing reliability, because it ensures that all the readers analyze the same images [8,33]. Agreement in every feature of the SINS was not analyzed, and some items have shown to lead to only poor to fair agreement [8,10,20], whereas others, such as vertebral osteolysis and kyphotic deformity, predict the occurrence of compression fracture after radiotherapy better than the whole SINS score [34–36]. However, the present study focused on the reliability of the global SINS score, which is the relevant feature for identifying patients eligible for surgery.

All patients underwent MRI and CT imaging. Computed tomography imaging is more accurate than radiography for depicting bone quality [37], and agreement in the SINS score might have been different if the latter had been used [10,20]. However, CT imaging is routinely used to assess spine metastatic disease within the SNHS and most Western countries. Readers were volunteers from each of the invited hospital departments and were not randomly selected. Therefore, selection bias may exist; it is possible that physicians who agreed to participate in this study were those who were the most motivated or interested in spine metastatic disease [38]. Should this be the case, agreement might be lower among other clinicians less familiar with spine metastatic disease, and it is impossible to completely rule out this possibility. Nevertheless, the number of participants was large, they came from different specialties and settings, and agreement was similar irrespective of the number of years of experience and across all types of hospitals [22]. All of the above suggests that results from the present study are valid in routine clinical practice.

In conclusion, the present study suggests that the agreement in the SINS score among radiologists, medical oncologists, radiation oncologists, orthopedic surgeons, and neurosurgeons is "moderate" and "almost perfect" when identifying the spine levels involved, which supports generalizing its use in routine clinical practice.

Acknowledgments

Thank you to Prof. David Moratal for his continuous collaboration.

References

[1] Coleman RE. Clinical features of metastatic bone disease and risk of skeletal morbidity. Clin Cancer Res 2006;12:6243s–6249s. doi:10.1158/1078-0432.CCR-06-0931.
[2] Weber MH, Burch S, Buckley J, Schmidt MH, Fehlings MG, Vrionis FD, et al. Instability and impending instability of the thoracolumbar spine in patients with spinal metastases: a systematic review. Int J Oncol 2011;38:5–12. doi:10.3892/ijo_00000818.
[3] Taneichi H, Kaneda K, Takeda N, Abumi K, Satoh S. Risk factors and probability of vertebral body collapse in metastases of the thoracic and lumbar spine. Spine (Phila Pa 1976) 1997;22:239–45.
[4] Izzo R, Guarnieri G, Guglielmi G, Muto M. Biomechanics of the spine. Part II: spinal instability. Eur J Radiol 2013;82:127–38. doi:10.1016/j.ejrad.2012.07.023.
[5] Kaloostian PE, Yurter A, Zadnik PL, Sciubba DM, Gokaslan ZL. Current paradigms for metastatic spinal disease: an evidence-based review. Ann Surg Oncol 2014;21:248–62. doi:10.1245/s10434-013-3324-8.
[6] Brooks FM, Ghatahora A, Brooks MC, Warren H, Price L, Brahmabhatt P, et al. Management of metastatic spinal cord compression: awareness of NICE guidance. Eur J Orthop Surg Traumatol 2014;24(Suppl. 1):S255–9. doi:10.1007/s00590-014-1438-8.
[7] Fisher CG, DiPaola CP, Ryken TC, Bilsky MH, Shaffrey CI, Berven SH, et al. A novel classification system for spinal instability in neoplastic disease: an evidence-based approach and expert consensus from the Spine Oncology Study Group. Spine (Phila Pa 1976) 2010;35:E1221–9. doi:10.1097/BRS.0b013e3181e16ae2.
[8] Fourney DR, Frangou EM, Ryken TC, Dipaola CP, Shaffrey CI, Berven SH, et al. Spinal Instability Neoplastic Score: an analysis of reliability and validity from the spine oncology study group. J Clin Oncol 2011;29:3072–7. doi:10.1200/JCO.2010.34.3897.
[9] Roberts CC, Daffner RH, Weissman BN, Bancroft L, Bennett DL, Blebea JS, et al. ACR appropriateness criteria on metastatic bone disease. J Am Coll Radiol 2010;7:400–9. doi:10.1016/j.jacr.2010.02.015.
[10] Campos M, Urrutia J, Zamora T, Román J, Canessa V, Borghero Y, et al. The Spine Instability Neoplastic Score: an independent reliability and reproducibility analysis. Spine J 2014;14:1466–9. doi:10.1016/j.spinee.2013.08.044.
[11] Huisman M, van der Velden JM, van Vulpen M, van den Bosch MAAJ, Chow E, Öner FC, et al. Spinal instability as defined by the Spinal Instability Neoplastic Score is associated with radiotherapy failure in metastatic spinal disease. Spine J 2014;14:2835–40. doi:10.1016/j.spinee.2014.03.043.
[12] Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011;64:96–106. doi:10.1016/j.jclinepi.2010.03.002.
[13] Departamento de Métodos Cuantitativos en Economía y Gestión U de LP de GC. Clasificación de hospitales públicos españoles mediante el uso del análisis cluster 2007. Available at: http://www.icmbd.es/docs/resumenClusterHospitales.pdf, accessed January 25, 2014.
[14] Fleiss JL. The Design and Analysis of Clinical Experiments. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1986. p. 1–32 doi:10.1002/9781118032923.
[15] Fleiss JL, Nee JC, Landis JR. Large sample variance of kappa in the case of different sets of raters. Psychol Bull 1979;86:974–7. doi:10.1037/0033-2909.86.5.974.
[16] Efron B. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. Biometrika 1981;68:589–99. doi:10.1093/biomet/68.3.589.
[17] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159. doi:10.2307/2529310.
[18] Keil S, Barabasch A, Dirrichs T, Bruners P, Hansen NL, Bieling HB, et al. Target lesion selection: an important factor causing variability of response classification in the Response Evaluation Criteria for Solid Tumors 1.1. Invest Radiol 2014;49:509–17. doi:10.1097/RLI.0000000000000048.
[19] Fisher CG, Versteeg AL, Schouten R, Boriani S, Varga PP, Rhines LD, et al. Reliability of the spinal instability neoplastic scale among radiologists: an assessment of instability secondary to spinal metastases. AJR Am J Roentgenol 2014;203:869–74. doi:10.2214/AJR.13.12269.
[20] Teixeira WGJ, Coutinho PDM, Marchese LD, Narazaki DK, Cristante AF, Teixeira MJ, et al. Interobserver agreement for the Spine Instability Neoplastic Score varies according to the experience of the evaluator. Clinics (Sao Paulo) 2013;68:213–18.
[21] Fisher CG, Schouten R, Versteeg AL, Boriani S, Varga PP, Rhines LD, et al. Reliability of the Spinal Instability Neoplastic Score (SINS) among radiation oncologists: an assessment of instability secondary to spinal metastases. Radiat Oncol 2014;9:69. doi:10.1186/1748-717X-9-69.
[22] Obuchowski NA. How many observers are needed in clinical studies of medical imaging? AJR Am J Roentgenol 2004;182:867–9. doi:10.2214/ajr.182.4.1820867.
[23] Brorson S, Hróbjartsson A. Training improves agreement among doctors using the Neer system for proximal humeral fractures in a systematic review. J Clin Epidemiol 2008;61:7–16. doi:10.1016/j.jclinepi.2007.04.014.
[24] Lee EH, Jun JK, Jung SE, Kim YM, Choi N. The efficacy of mammography boot cAMP to improve the performance of radiologists. Korean J Radiol 2014;15:578–85. doi:10.3348/kjr.2014.15.5.578.
[25] Bankier AA, Levine D, Halpern EF, Kressel HY. Consensus interpretation in imaging research: is there a better way? Radiology 2010;257:14–17. doi:10.1148/radiol.10100252.
[26] Dea N, Fisher CG. Evidence-based medicine in metastatic spine disease. Neurol Res 2014;36:524–9. doi:10.1179/1743132814Y.0000000365.
[27] Prasad TV, Thulkar S, Hari S, Sharma DN, Kumar S. Role of computed tomography (CT) scan in staging of cervical carcinoma. Indian J Med Res 2014;139:714–19.
[28] Tsuda H, Akiyama F, Kurosumi M, Sakamoto G, Watanabe T. The efficacy and limitations of repeated slide conferences for improving interobserver agreement when judging nuclear atypia of breast cancer. The Japan National Surgical Adjuvant Study of Breast Cancer (NSAS-BC) Pathology Section. Jpn J Clin Oncol 1999;29:68–73.
[29] Keating NL, Landrum MB, Lamont EB, Bozeman SR, Shulman LN, McNeil BJ. Tumor boards and the quality of cancer care. J Natl Cancer Inst 2013;105:113–21. doi:10.1093/jnci/djs502.
[30] Prades J, Remue E, van Hoof E, Borras JM. Is it worth reorganising cancer services on the basis of multidisciplinary teams (MDTs)? A systematic review of the objectives and organisation of MDTs and their impact on patient outcomes. Health Policy (New York) 2014;doi:10.1016/j.healthpol.2014.09.006 (in press).
[31] Sahgal A, Fehlings MG. In reply to Fourney. Int J Radiat Oncol Biol Phys 2013;85:894–5. doi:10.1016/j.ijrobp.2012.10.018.
[32] Schlampp I, Rieken S, Habermehl D, Bruckner T, Förster R, Debus J, et al. Stability of spinal bone metastases in breast cancer after radiotherapy: a retrospective analysis of 157 cases. Strahlenther Onkol 2014;190:792–7. doi:10.1007/s00066-014-0651-z.
[33] Dimopoulos JCA, De Vos V, Berger D, Petric P, Dumas I, Kirisits C, et al. Inter-observer comparison of target delineation for MRI-assisted

cervical cancer brachytherapy: application of the GYN GEC-ESTRO recommendations. Radiother Oncol 2009;91:166–72. doi:10.1016/j.radonc.2008.10.023.
[34] Sung S-H, Chang U-K. Evaluation of risk factors for vertebral compression fracture after stereotactic radiosurgery in spinal tumor patients. Korean J Spine 2014;11:103–8. doi:10.14245/kjs.2014.11.3.103.
[35] Cunha MVR, Al-Omair A, Atenafu EG, Masucci GL, Letourneau D, Korol R, et al. Vertebral compression fracture (VCF) after spine stereotactic body radiation therapy (SBRT): analysis of predictive factors. Int J Radiat Oncol Biol Phys 2012;84:e343–9. doi:10.1016/j.ijrobp.2012.04.034.
[36] Sahgal A, Atenafu EG, Chao S, Al-Omair A, Boehling N, Balagamwala EH, et al. Vertebral compression fracture after spine stereotactic body radiotherapy: a multi-institutional analysis with a focus on radiation dose and the Spinal Instability Neoplastic Score. J Clin Oncol 2013;31:3426–31. doi:10.1200/JCO.2013.50.1411.
[37] Hamaoka T, Madewell JE, Podoloff DA, Hortobagyi GN, Ueno NT. Bone imaging in metastatic breast cancer. J Clin Oncol 2004;22:2942–53. doi:10.1200/JCO.2004.08.181.
[38] Lim C, Cheung MC, Franco B, Dharmakulaseelan L, Chong E, Iyngarathasan A, et al. Quality improvement: an assessment of participation and attitudes of medical oncologists. J Oncol Pract 2014;10:e408–14. doi:10.1200/JOP.2014.001515.

4. Fracturas vertebrales metastásicas (FVM) frente osteoporóticas en RM; una evaluación de acuerdo observador ciega, multicéntrica y multiespecialidad.

J Natl Compr Canc Netw 2020;18(3):1–8

Antecedentes: Se asume que la Resonancia Magnética (RM) es válida para diferenciar fracturas vertebrales metastásicas (FVM) de las fracturas osteoporóticas (FVO). Este estudio comprobó (1) la concordancia entre el diagnóstico basado en la imagen de FVM frente FVO y el patrón de referencia (biopsia o seguimiento mayor de 6 meses), (2) el acuerdo intra- e interobservador en las hallazgos claves de imagen y el diagnóstico de FVM vs FVO, y (3) si mostrar el dato de la historia del paciente respecto al cáncer lleva a variaciones en el diagnóstico, concordancia o acuerdo.

Pacientes y Métodos: Este estudio retrospectivo de cohorte incluyó los datos clínicos y de imagen de 203 pacientes con FVM o FVO confirmada a 25 médicos (neurocirujanos, radiólogos, cirujanos ortopédicos y oncólogos radioterapeutas). Desde enero de 2018 a octubre de 2018, los especialistas interpretaron las imágenes en condiciones lo más cercanas posible a la práctica clínica. Cada especialista estudió los datos dos veces, con un intervalo mínimo de 6 semanas, ciego a los resultados de otros clínicos o a los suyos propios. El estadístico kappa se empleó para comprobar el acuerdo intra- e interobservador en los hallazgos claves de imagen, diagnóstico (FVM vs FVO) y concordancia con el patrón de referencia. Los análisis de subgrupos se basaron en la especialidad clínica, años de experiencia y complejidad del hospital donde trabajaban.

Resultados: Para el diagnóstico de FVM vs FVO, el acuerdo interobservador fue discreto, mientras el acuerdo intraobservador fue sustancial. Sólo esta última mejoró a casi perfecta con se mostró los antecedentes de cáncer del paciente. El acuerdo interobservador para los hallazgos de imagen clave fue discreto o moderado, mientras el acuerdo intraobservador fue moderado o sustancial. La concordancia entre el diagnóstico de FVM vs FVO y la referencia fue moderada. Los resultados fueron similares independientemente de la especialidad del médico, experiencia o categoría del hospital.

Conclusiones: Con la RM para diagnóstica FVM vs FVO, el acuerdo interobservador y la concordancia con el patrón de referencia fueron moderados. Estos resultados muestran las dudas en la reproducibilidad de basar estos diagnóstico en la práctica clínica con la RM.

Metastatic Versus Osteoporotic Vertebral Fractures on MRI: A Blinded, Multicenter, and Multispecialty Observer Agreement Evaluation

Estanislao Arana, MD, MHE, PhD[a,b]; Francisco M. Kovacs, MD, PhD[b,c]; Ana Royuela, PhD[b,d]; Beatriz Asenjo, MD, PhD[b,e]; Fatima Nagib, MD[b,e]; Sandra Pérez-Aguilera, MD[b,f]; María Dejoz, BEng[b,g]; Alberto Cabrera-Zubizarreta, MD[b,h]; Yolanda García-Hidalgo, MD, PhD[b,i]; and Ana Estremera, MD, PhD[b,j]; for the Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis*

ABSTRACT

Background: MRI is assumed to be valid for distinguishing metastatic vertebral fractures (MVFs) from osteoporotic vertebral fractures (OVFs). This study assessed (1) concordance between the image-based diagnosis of MVF versus OVF and the reference (biopsy or follow-up of >6 months), (2) interobserver and intraobserver agreement on key imaging findings and the diagnosis of MVF versus OVF, and (3) whether disclosing a patient's history of cancer leads to variations in diagnosis, concordance, or agreement. **Patients and Methods:** This retrospective cohort study included clinical data and imaging from 203 patients with confirmed MVF or OVF provided to 25 clinicians (neurosurgeons, radiologists, orthopedic surgeons, and radiation oncologists). From January 2018 through October 2018, the clinicians interpreted images in conditions as close as possible to routine practice. Each specialist assessed data twice, with a minimum 6-week interval, blinded to assessments made by other clinicians and to their own previous assessments. The kappa statistic was used to assess interobserver and intraobserver agreement on key imaging findings, diagnosis (MVF vs OVF), and concordance with the reference. Subgroup analyses were based on clinicians' specialty, years of experience, and complexity of the hospital where they worked. **Results:** For diagnosis of MVF versus OVF, interobserver agreement was fair, whereas intraobserver agreement was substantial. Only the latter improved to almost perfect when a patient's history of cancer was disclosed. Interobserver agreement for key imaging findings was fair or moderate, whereas intraobserver agreement on key imaging findings was moderate or substantial. Concordance between the diagnosis of MVF versus OVF and the reference was moderate. Results were similar regardless of clinicians' specialty, experience, and hospital category. **Conclusions:** When MRI is used to distinguish MVF versus OVF, interobserver agreement and concordance with the reference were moderate. These results cast doubt on the reliability of basing such a diagnosis on MRI in routine practice.

J Natl Compr Canc Netw 2020;18(3):267–273
doi: 10.6004/jnccn.2019.7367

Background

Nontraumatic vertebral fractures are frequently seen in clinical practice. Most are caused by osteoporosis and are diagnosed as osteoporotic vertebral fractures (OVFs), but metastatic vertebral fractures (MVFs) are also common. Determining whether a vertebral fracture has been caused by MVF or OVF is key for establishing appropriate treatment and prognosis and can have a profound psychological impact on patients. Therefore, the accuracy and reliability of the data used to reach this diagnosis are paramount.

Several imaging findings are frequently used to help distinguish between OVF and MVF.[1,2] Some have been fed into risk-scoring algorithms developed to identify patients at a higher risk of experiencing MVF.[3] To be useful in clinical practice and lead to sound treatment decisions, risk-assessment algorithms should be evidence-based and built on parameters that can be assessed reliably. However, the available risk-scoring algorithms in this field rely on ancillary imaging findings, for which

[a]Department of Radiology, Fundación Instituto Valenciano de Oncología, Valencia; [b]Spanish Back Pain Research Network, Kovacs Foundation, Palma de Mallorca; [c]Unidad de la Espalda Kovacs, Hospital Universitario HLA-Moncloa, Madrid; [d]Clinical Biostatistics Unit, Instituto de Investigación Sanitaria Puerta de Hierro-Segovia de Arana, Madrid; [e]Department of Radiology, Hospital Universitario Regional de Málaga, Málaga; [f]Department of Radiology, Hospital de Manacor, Mallorca; [g]School of Biomedical Engineering, Universitat Politècnica de Valencia, Valencia; [h]Department of Radiology, Hospital de Galdakao, Galdakao, Bizkaia; [i]Department of Radiology, Hospital Universitario Puerta de Hierro, Madrid; and [j]Department of Radiology, Hospital Son Llàtzer, Palma de Mallorca, Spain.

*To view additional members of the Spanish Back Pain Research Network Task Force, see supplemental eAppendix 1 (available with this article at JNCCN.org).

See JNCCN.org for supplemental online content.

See page 362 for related commentary.

reliability is unknown. The need to assess their reliability has been previously highlighted.[3,4]

The available scoring systems have been developed based on the interpretation of images by only one observer[1] or on the consensus of readers working in the same institution who tested the validity of their scoring systems with a small number of patients.[2,3] However, in clinical practice, when patients seek care for back pain caused by a nontraumatic vertebral fracture, spine imaging can be assessed by practitioners from an array of specialties, and management of OVF and especially MVF is multidisciplinary.

Therefore, the purpose of this study was to assess among clinicians from different specialties and working in different healthcare centers, in conditions as close as possible to routine clinical practice, (1) concordance between the clinical diagnosis (MVF vs OVF) and the reference (diagnosis established by biopsy or clinical follow-up), (2) interobserver and intraobserver agreement on the diagnosis of MVF versus OVF and on the interpretation of key imaging findings leading to such diagnosis, and (3) whether concordance and agreements improve when clinicians are aware of a patient's history of cancer.

Patients and Methods

This study was approved by the Institutional Review Boards of the participating hospitals and complied with the Guidelines for Reporting Reliability and Agreement Studies.[5] Written informed consent was waived because of the retrospective nature of this study.

Setting and Participants

Patients and images were selected by a radiologist with 25 years of experience who did not participate in image interpretation. He revised records from his hospital in reverse chronologic order (ie, more recent cases were revised first) and selected cases complying with the inclusion criteria until the sample size was reached. The radiologist then selected 3 images per patient: 2 sagittal images on T1-, T2-, or short inversion time inversion-recovery (STIR)–weighted images, and 1 axial T1-weighted image.

Inclusion criteria were having requested care for a nontraumatic vertebral fracture, and diagnosis of MVF or OVF confirmed through biopsy or clinical follow-up of >6 months. Exclusion criteria were missing clinical history for any of the data required by the readers, and imaging of insufficient quality to assess the spinal levels affected (Figure 1).

A total of 22 hospital departments of radiology, radiation oncology, orthopedic surgery, and neurosurgery were invited to join the study because they had participated in previous spine studies undertaken by the Spanish Back Pain Research Network or had expressed interest in doing so. The hospital departments were located in 18 hospitals across 12 geographic regions; 6 departments were located in 5 private hospitals

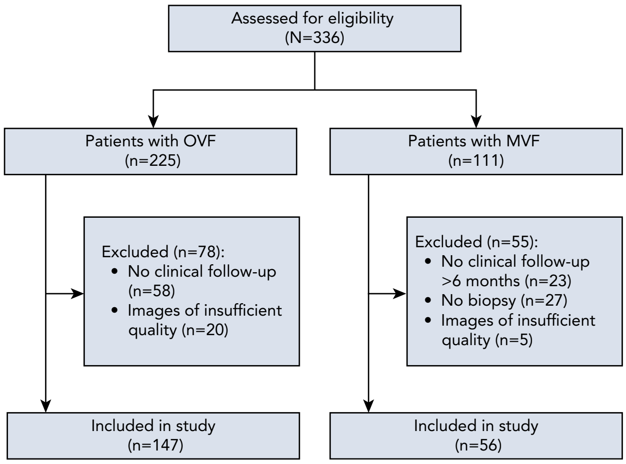


Figure 1. Flowchart of the selection process.
Abbreviations: MVF, metastatic vertebral fracture; OVF, osteoporotic vertebral fracture.

and 16 were located in 13 nonprofit hospitals belonging to or working for the Spanish National Health System (SNHS). The SNHS is the tax-funded, government-owned organization that provides free healthcare to every resident in Spain. The SNHS classifies hospitals into 5 categories based on their complexity,[6] with category 1 the simplest and category 5 the most complex. Departments invited to participate in this study were located in category 2 through 5 hospitals.

According to standard procedure in our setting, neither subjects nor clinicians received any compensation for their involvement in this study.

MRI Evaluation, Reporting, and Interpretation

All images were acquired on 4 1.5T MRI systems, using similar sequences (supplemental eTable 1, available with this article at JNCCN.org).

The recruiting radiologist prepared an information pack on each patient containing 3 images and a clinical vignette summarizing the patient's age, oncologic history, and clinical signs and symptoms.[7] Patient identity was masked and a code was assigned to each pack. All packs were uploaded to an online platform designed for this study (http://www.typeform.com/). The 3 images included 2 sagittal images on T1-, T2-, or STIR-weighted images and 1 axial T1-weighted image. The radiologist segmented the selected images so that readers were shown the index vertebral segment, the one immediately above, and the one immediately below. In the case of patients showing vertebral fractures at several levels, the radiologist defined the index as the one showing a recent fracture, at the level for which the patient had requested care, and that was subject to biopsy or clinical follow-up for >6 months.

MRI findings assessed in this study were selected through a literature review[1,2] and are shown in supplemental eTable 2. They include those findings used to calculate the MRI Evaluation Totalizing Assessment (META) score.[1] The readers assessed all MRI images on their own, prospectively, from January 2018 through October 2018, using an in-house online MRI interpretation system. No attempt was made to homogenize their diagnostic criteria or interpretation of images. Readers were told to use their own clinical judgment as they would in routine clinical practice and to upload the report directly onto the online platform. After they assessed the imaging findings, readers were requested to state their diagnosis ("MVF" vs "OVF"). Finally, after the patient's cancer history was disclosed, readers were given the opportunity to modify their diagnosis (Figure 2), and modifications were recorded.

Readers assessed each information pack twice, with a minimum 6-week interval between the 2 rounds. After the information from the first round was uploaded, the platform software made it impossible for readers to access it again until the interval had elapsed. It also denied access to colleagues' reports and to their own previous reports.

Data introduced into the platform were automatically converted into a spreadsheet. The software engineer in charge of developing the platform cross-checked to ensure that data in the database matched the information that readers had introduced into the platform.

Statistical Analysis

To assess interobserver and intraobserver agreement, ratings from each observer were cross-tabulated, and agreement was measured using the kappa statistic (κ) with the corresponding 95% confidence interval for interobserver agreement and the percentiles 25 and 75 (interquartile range [IQR], p25–p75) for intraobserver agreement. Kappa values were categorized as reflecting an "almost perfect" (0.81–1.00), "substantial" (0.61–0.80), "moderate" (0.41–0.60), "fair" (0.21–0.40), "slight" (0.00–0.20), or "poor" ($<$0.00) agreement.[8]

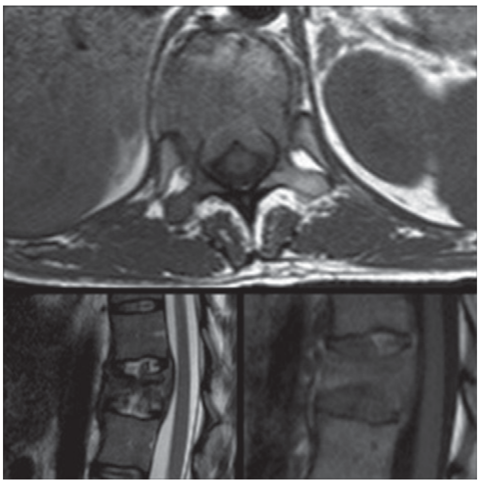
The association between the diagnosis (MVF vs OVF) before and after readers were aware of a patient's cancer history was based on the assessments made during the first round. Diagnostic accuracy was defined as the concordance between each reader's diagnosis at the first round (MVF vs OVF) and the reference diagnosis. Concordance was measured using the kappa statistic. In a subgroup analysis, diagnostic accuracy was measured separately for subjects who presented and did not present previous fractures on imaging.

Sample size was estimated at 203 patients with vertebral fractures, assuming that (1) vertebral fractures would be caused by MVF in 25% to 30% of cases,[9] (2) the minimal number of assessments to be compared would be 2 (for intraobserver agreement), and (3) the kappa index would be $\geq$0.7 with a confidence margin of 0.10 on each side.

Results

All 22 hospital departments invited to join the study accepted, and 25 clinicians from these departments participated: 9 radiologists, 4 radiation oncologists, 5 orthopedic surgeons, and 7 neurosurgeons (Table 1). The number of years (after residency) that the clinicians had been interpreting spine MRIs in routine practice on a daily basis ranged from 4 to 35 years. Table 1 also shows the characteristics of the 203 patients whose clinical histories and images were selected for the study and of the 25 readers who interpreted their data.

Case 1. Female, age 34 y



- Normal vertebral signal replace with bone marrow edema
  - ☐ Pratially or completely
  - ☐ Showing a bandlike pattern
- Horizontal frecture line on fluid-sensitive sequence (STIR) or T2-weighted imaging: ☐Yes / ☐No
- Deposit-like appearance of pedicle involvement: ☐Yes / ☐No
- Convexity of the posterior vertebral body border (bulging posterior cortex): ☐Yes / ☐No
- Posterosuperior retropulsion: ☐Yes / ☐No
- Symmetry of the signal intensity changes: ☐Symmetric ☐Asymmetric
- Diagnosis: ☐VFO ☐MVF
- This patient has a history of cancer: ☐Yes / ☐No
- Do you want to modify your diagnosis? ☐Yes / ☐No
- Diagnosis: ☐OVF ☐MVF

Figure 2. Sample imaging finding.
Abbreviations: MVF, metastatic vertebral fracture; OVF, osteoporotic vertebral fracture; STIR, short inversion time inversion-recovery.

Table 1. Sample Characteristics

Characteristic	n (%)
Patients	
Total, N	203
Age, mean (SD), y	60.8 (12.3)
Women	62.1 (14.5)
Men	61.7 (10.8)
Sex (female)	139 (68.47)
Location of spinal fracture	
Thoracic	98 (48.27)
Lumbar	105 (51.73)
Diagnosis[a]	
OVF	147 (72.4)
MVF	56 (27.6)
History of cancer	
No	122 (60.1)
Yes	81 (39.9)
Previous spine fractures	
No	132 (65)
Yes	71 (35)
Location of previous spine fracture	
Thoracic	36 (50.7)
Lumbar	35 (49.3)
Primary malignancies	
Lung	20 (35.7)
Breast	16 (28.5)
Colon	8 (14.3)
Lymphoma or myeloma	4 (7.1)
Other	8 (14.3)
Readers	
Total, N	25
Specialty	
Radiology	9 (36.0)
Radiation oncology	4 (16.0)
Orthopedic surgery	5 (20.0)
Neurosurgery	7 (28.0)
Years of experience (postresidency), y	
≤7	7 (32.5)
8–13	7 (30.1)
≥14	11 (37.4)
Hospital category (complexity)[b]	
2	3 (3.6)
3	7 (30.1)
4	7 (22.9)
5 (most complex)	8 (43.4)

(continued)

Table 1. Sample Characteristics (cont.)

Characteristic	n (%)
Hospitals	
Total, N	18
Management	
Nonprofit[c]	13 (72.2)
For-profit[d]	5 (27.8)
Departments, N	22
Radiology	8 (40.0)
Radiation oncology	4 (20.0)
Orthopedic surgery	3 (15.0)
Neurosurgery	7 (25.0)

Abbreviations: MVF, metastatic vertebral fracture; OVF, osteoporotic vertebral fracture; SNHS, Spanish National Health System.
[a]Diagnosis established by the reference (biopsy or follow-up >6 months).
[b]Based on size, availability of high-tech medical equipment and procedures, and degree of educational activity. No readers from category 1 hospitals (simplest) were included in this study.
[c]Belonging to and managed by the SNHS, or belonging to or managed by charities working for the SNHS.
[d]Privately owned and managed.

As Table 2 shows, interobserver agreement in the diagnosis of MVF versus OVF was fair (κ, 0.397; 95% CI, 0.347–0.450) when the reader was unaware of the patient's history of cancer. When the patient's history of cancer was disclosed, the agreement increased to moderate (κ, 0.467; 95% CI, 0.418–0.518).

Intraobserver agreement on the diagnosis of MVF versus OVF was substantial (κ, 0.624; IQR, 0.517–0.693), and improved to almost perfect after the patient's history of cancer was disclosed (κ, 0.878; IQR, 0.781–0.939 and κ, 0.851; IQR, 0.779–0.948 at the first and second rounds, respectively). This increase in agreement was observed across all clinical specialties, with orthopedic surgery showing the highest increase (from κ, 0.588; IQR, 0.509–0.595 to κ, 0.917; IQR, 0.859–0.959) (Table 3).

Interobserver agreement was moderate on "deposit-like appearance of pedicle involvement" and "bulging posterior cortex" and fair on all the other imaging findings (supplemental eTable 3). Agreement among radiologists was moderate for most imaging findings, but no consistent differences were found among clinical specialties (supplemental eTable 3).

Intraobserver agreement on individual imaging findings ranged from moderate to substantial and was similar across clinical specialties (supplemental eTable 4).

After being informed of a patient's clinical history of cancer, the readers modified the diagnosis (MVF vs OVF) of 142 patients (69.5%). All the readers modified the diagnosis of at least 1 patient (range of number of patients for whom each clinician changed the diagnosis, 1–39). Among the 5,075 assessments made by the 25 readers using the 203 images, the previous diagnosis was changed in

Table 2. Interobserver Agreement

	Kappa (95% CI)
All readers (n=25)	
Diagnosis of OVF vs MVF (before disclosing history of cancer)	0.397 (0.347–0.450)
Pattern of signal abnormalities	0.396 (0.349–0.445)
Horizontal fracture line	0.220 (0.177–0.266)
Deposit-like appearance of pedicle involvement	0.447 (0.395–0.501)
Bulging posterior cortex	0.426 (0.383–0.472)
Posterosuperior retropulsion	0.319 (0.280–0.359)
Symmetry of signal intensity changes	0.270 (0.230–0.312)
Diagnosis of OVF vs MVF (after disclosing history of cancer)	0.467 (0.418–0.518)
Radiology (n=9)	
Diagnosis of OVF vs MVF (before disclosing history of cancer)	0.508 (0.446–0.573)
Diagnosis of OVF vs MVF (after disclosing history of cancer)	0.574 (0.518–0.633)
Neurosurgery (n=7)	
Diagnosis of OVF vs MVF (before disclosing history of cancer)	0.364 (0.305–0.425)
Diagnosis of OVF vs MVF (after disclosing history of cancer)	0.456 (0.397–0.518)
Orthopedic surgery (n=5)	
Diagnosis of OVF vs MVF (before disclosing history of cancer)	0.342 (0.275–0.411)
Diagnosis of OVF vs MVF (after disclosing history of cancer)	0.370 (0.303–0.438)
Radiation oncology (n=4)	
Diagnosis of OVF vs MVF (before disclosing history of cancer)	0.321 (0.256–0.389)
Diagnosis of OVF vs MVF (after disclosing history of cancer)	0.394 (0.325–0.465)

Abbreviations: MVF, metastatic vertebral fracture; OVF, osteoporotic vertebral fracture.

5.0% of the patients without a history of cancer versus 10.8% of those with a history of cancer (chi-square, $P<.001$).

Before readers were aware of a patient's clinical history of cancer, concordance of their diagnosis with the reference was moderate (κ, 0.437; IQR, 0.326–0.511). Having access to a patient's history only marginally improved concordance (κ, 0.443; IQR, 0.398–0.526). Diagnostic accuracy was only fair for orthopedic surgeons, whereas it was moderate for all other specialties. However, differences in κ values were minimal, and the IQR values overlapped. Diagnostic accuracy was very similar regardless of years of professional experience and category of hospital (supplemental eTable 5).

Concordance with the reference for subjects without images of preexisting fractures was κ=0.452 (IQR, 0.387–0.509) before the clinical history of cancer was disclosed and κ=0.462 (IQR, 0.407–0.570) after it was disclosed. For subjects with preexisting fractures, these values were κ=0.286 (IQR, 0.183–0.396) and κ=0.331 (IQR, 0.219–0.368), respectively (supplemental eTable 6).

Discussion

In routine practice, the suspicion of MVF or OVF is based on clinical history and imaging. Our findings showed that interobserver agreement was fair and that diagnostic accuracy was moderate.

This is the first study to analyze the reliability of the diagnosis of MVF versus OVF using a large multidisciplinary team of readers working in different healthcare centers and assessing diagnostic accuracy against a reference. It was conducted in conditions as close as possible to routine clinical practice; readers were provided with actual clinical histories.[10] Because no instructions, scoring systems, or meetings were implemented to improve agreement,[11–13] clinicians had to make their diagnosis on their own based on data from clinical history and imaging, with common heuristics and biases.[14] All of these factors may account for differences between the results of this study and the almost perfect agreement reported by the medical professionals who developed the META score (κ, 0.93),[1] which previous studies have shown to not be reproducible.[15]

In this study, readers were experts who had been managing vertebral fractures and interpreting spine imaging for up to 35 years, had participated in previous research in this field, and felt confident enough to volunteer for a study assessing their interpretation of spine images. Diagnostic accuracy was very similar across clinical specialties, readers' experience, and hospital category and was consistent with results from the few previous studies that analyzed the reproducibility of single imaging findings and the META score.[1,15] Therefore, fair interobserver agreement and moderate diagnostic accuracy may be the best that can be realistically expected when using MRI to distinguish MVF versus OVF in routine practice, simply because with current technology, images of MVF and OVF are sometimes indistinguishable.[16,17] For instance, "bulging posterior cortex" was one of the imaging findings with the best interobserver agreement found in this and previous studies, and specifically, expansion of the posterior aspect of the vertebral contour is associated with malignant fractures.[18] However, it can also be observed in benign OVFs, especially in acute posttraumatic fractures.[16]

The low reproducibility of imaging findings challenges the validity of purportedly evidence-based decision support systems based on them.[2] In fact, a decision system based on unreliable findings can be detrimental.[3] The degree of agreement found in this and previous studies would classify MRI as class II for diagnosing MVF versus OVF and as class III for assessing individual imaging findings.[19]

In general, disclosing accurate clinical data slightly increases the accuracy of diagnostic tests.[20] For imaging

Table 3. Intraobserver Agreement

	Median Kappa (IQR)
All readers (n=25)	
Diagnosis of OVF vs MVF (agreement between diagnosis in both rounds, before disclosing history of cancer)[a]	0.624 (0.517–0.693)
Pattern of signal abnormalities	0.660 (0.555–0.762)
Horizontal fracture line	0.535 (0.457–0.683)
Deposit-like appearance of pedicle involvement	0.653 (0.549–0.732)
Bulging posterior cortex	0.715 (0.618–0.824)
Posterosuperior retropulsion	0.673 (0.592–0.731)
Symmetry of signal intensity changes	0.489 (0.402–0.646)
Diagnosis of OVF vs MVF (agreement before and after disclosing history of cancer), first round[b]	0.878 (0.781–0.939)
Diagnosis of OVF vs MVF (agreement before and after disclosing history of cancer), second round[c]	0.851 (0.779–0.948)
Radiology (n=9)	
Diagnosis of OVF vs MVF (agreement between diagnosis in both rounds, before disclosing history of cancer)[a]	0.652 (0.630–0.733)
Diagnosis of OVF vs MVF (agreement before and after disclosing history of cancer), first round[b]	0.867 (0.805–0.881)
Diagnosis of OVF vs MVF (agreement before and after disclosing history of cancer), second round[c]	0.845 (0.779–0.870)
Neurosurgery (n=7)	
Diagnosis of OVF vs MVF (agreement between diagnosis in both rounds, before disclosing history of cancer)[a]	0.550 (0.483–0.693)
Diagnosis of OVF vs MVF (agreement before and after disclosing history of cancer), first round[b]	0.877 (0.713–0.979)
Diagnosis of OVF vs MVF (agreement before and after disclosing history of cancer), second round[c]	0.948 (0.832–1.000)

(continued)

Table 3. Intraobserver Agreement (cont.)

	Median Kappa (IQR)
Orthopedic surgery (n=5)	
Diagnosis of OVF vs MVF (agreement between diagnosis in both rounds, before disclosing history of cancer)[a]	0.588 (0.509–0.595)
Diagnosis of OVF vs MVF (agreement before and after disclosing history of cancer), first round[b]	0.917 (0.859–0.959)
Diagnosis of OVF vs MVF (agreement before and after disclosing history of cancer), second round[c]	0.880 (0.871–0.930)
Radiation oncology (n=4)	
Diagnosis of OVF vs MVF (agreement between diagnosis in both rounds, before disclosing history of cancer)[a]	0.618 (0.575–0.683)
Diagnosis of OVF vs MVF (agreement before and after disclosing history of cancer), first round[b]	0.912 (0.706–0.958)
Diagnosis of OVF vs MVF (agreement before and after disclosing history of cancer), second round[c]	0.761 (0.581–0.921)

Abbreviations: IQR, interquartile range; MVF, metastatic vertebral fracture; OVF, osteoporotic vertebral fracture.
[a]This kappa value reflects the agreement between the diagnosis established by the same reader based on the same images, at the first and second rounds (in both cases, before being aware of patient's history of cancer).
[b]This kappa value reflects the agreement between the diagnosis established by the same reader based on the same images at the first round, before and after being aware of patient's cancer history.
[c]This kappa value reflects the agreement between the diagnosis established by the same reader based on the same images at the second round, before and after being aware of patient's cancer history.

procedures, some studies have suggested that accurate data disclosure decreases the interpretative performance,[21] whereas others have denied any negative consequences.[22] In our study, disclosing a patient's cancer history had no significant impact on interobserver agreement or diagnostic accuracy, but increased intraobserver agreement significantly and led to changes in the diagnosis of MVF versus OVF in 69.5% of the cases.

Diagnostic performance was similar across specialties. This is consistent with previous studies on the interpretation of spine imaging.[23–26] For patients with metastatic spine disease, surgeons' assessment of imaging is often considered the reference for referral to surgery.[27] However, no significant differences existed across surgical and nonsurgical specialties when these clinicians assessed the spinal instability score.[28] No patient was excluded due to sclerotic metastases, previous trauma history, or myeloma, in which vertebral signal intensity changes are misleading.[16,17] Interobserver agreement and diagnostic accuracy may be different for patients showing these findings.

This study has several limitations. The cases analyzed were selected by a radiologist and were not a random sample. These conditions were necessary to select a sample with the desired proportion of cases with MVF confirmed by a reference and is common practice in agreement studies on imaging or concordance.[13,26] In this study, readers only assessed 3 images, whereas in clinical practice physicians review multiple images. This rule was decided at the design phase of the study to enhance participation. Moreover, it is common practice in agreement studies to restrict the number of images to the most relevant or potentially confounding ones.[12,28] None of the selected cases showed findings highly suggestive of malignancy, such as soft tissue mass, which commonly lead to higher agreement between orthopedic surgeons and radiologists.[29] Therefore, it is possible that agreement would have been higher if a number of patients included in the study had shown these findings. However, this study aimed to assess agreement in conditions as close as possible to clinical practice, and inclusion criteria did not require any specific finding. The classification of imaging findings did not follow the categories established by the META score. This condition was decided at the design phase of the study because these categories have been shown to be unreliable.[15] Using different image sequences may lead to different results. However, MR imaging sequences are not widely available,[17] and were therefore considered inappropriate for a study replicating routine practice as closely as possible. Nevertheless, future studies should explore the impact of different image sequences on agreement and diagnostic accuracy.[17]

Conclusions

Diagnostic accuracy and interobserver agreement on the assessment of OVF versus MVF is moderate at best, irrespective of medical or surgical specialty, years of clinical experience, or hospital type. This result casts doubt on the reliability of using MRI findings together with clinical history as the basis for distinguishing OVF from MVF in routine clinical practice or multicenter studies.

Submitted July 7, 2019; accepted for publication October 7, 2019.

Author contributions: *Study concept:* Arana, Kovacs. *Study design:* Arana, Kovacs, Royuela. *Data collection:* Asenjo, Nagib, Pérez-Aguilera, Dejoz, Cabrera-Zubizarreta, García-Hidalgo, Estremera. *Quality assessment and logistics:* Arana, Dejoz. *Data analysis:* Royuela. *Guidance on clinical practice of vertebral fractures:* Arana. *Writing—original draft and revision:* Arana, Kovacs, Royuela. *Writing—review and editing:* All authors. *Critical revisions:* All authors.

Disclosures: The authors have disclosed that they have not received any financial consideration from any person or organization to support the preparation, analysis, results, or discussion of this article.

Funding: Funding for this work was provided by The Kovacs Foundation.

Correspondence: Estanislao Arana, MD, MHE, PhD, Servicio de Radiología, Fundación Instituto Valenciano de Oncología, C/ Beltrán Báguena, 19, 46009 Valencia, Spain. Email: Estanis.Arana@ext.uv.es

References

1. Kato S, Hozumi T, Yamakawa K, et al. META: an MRI-based scoring system differentiating metastatic from osteoporotic vertebral fractures. Spine J 2015;15:1563–1570.
2. Thawait SK, Kim J, Klufas RA, et al. Comparison of four prediction models to discriminate benign from malignant vertebral compression fractures according to MRI feature analysis. AJR Am J Roentgenol 2013;200:493–502.
3. Wang KC, Jeanmenne A, Weber GM, et al. An online evidence-based decision support system for distinguishing benign from malignant vertebral compression fractures by magnetic resonance imaging feature analysis. J Digit Imaging 2011;24:507–515.
4. Thawait SK, Marcus MA, Morrison WB, et al. Research synthesis: what is the diagnostic performance of magnetic resonance imaging to discriminate benign from malignant vertebral compression fractures? Systematic review and meta-analysis. Spine (Phila Pa 1976) 2012;37:E736–744.
5. Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011; 64:96–106.
6. Departamento de Métodos Cuantitativos en Economía y Gestión U de LP de GC. Clasificación de hospitales públicos españoles mediante el uso del análisis cluster [Internet], 2007. Accessed October 10, 2019. Available at: http://www.icmbd.es/docs/resumenClusterHospitales.pdf.
7. Bilsky MH, Laufer I, Fourney DR, et al. Reliability analysis of the epidural spinal cord compression scale. J Neurosurg Spine 2010;13:324–328.
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.
9. Curtis JR, Taylor AJ, Matthews RS, et al. "Pathologic" fractures: should these be included in epidemiologic studies of osteoporotic fractures? Osteoporos Int 2009;20:1969–1972.
10. Shankar PR, Kaza RK, Al-Hawary MM, et al. Impact of clinical history on maximum PI-RADS version 2 score: a six-reader 120-case sham history retrospective evaluation. Radiology 2018;288:158–163.
11. Levine D, Bankier AA, Halpern EF. Submissions to radiology: our top 10 list of statistical errors. Radiology 2009;253:288–290.
12. Elmore JG, Longton GM, Carney PA, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA 2015; 313:1122–1132.
13. Sherbeck JP, Zhao L, Lieberman RW. High variability in lymph node counts among an international cohort of pathologists: questioning the scientific validity of node counts. J Natl Compr Canc Netw 2018;16:395–401.
14. Itri JN, Patel SH. Heuristics and cognitive error in medical imaging. AJR Am J Roentgenol 2018;210:1097–1105.
15. Urrutia J, Besa P, Morales S, et al. Does the META score evaluating osteoporotic and metastatic vertebral fractures have enough agreement to be used by orthopaedic surgeons with different levels of training? Eur Spine J 2018;27:2577–2583.
16. Mauch JT, Carr CM, Cloft H, et al. Review of the imaging features of benign osteoporotic and malignant vertebral compression fractures. AJNR Am J Neuroradiol 2018;39:1584–1592.
17. Sung JK, Jee WH, Jung JY, et al. Differentiation of acute osteoporotic and malignant compression fractures of the spine: use of additive qualitative and quantitative axial diffusion-weighted MR imaging to conventional MR imaging at 3.0 T. Radiology 2014;271:488–498.
18. Besa P, Urrutia J, Campos M, et al. The META score for differentiating metastatic from osteoporotic vertebral fractures: an independent agreement assessment. Spine J 2018;18:2074–2080.
19. Shank CD, Lepard JR, Walters BC, et al. Towards evidence-based guidelines in neurological surgery. Neurosurgery 2019;85:613–621.
20. Loy CT, Irwig L. Accuracy of diagnostic tests read with and without clinical information: a systematic review. JAMA 2004;292:1602–1609.
21. Carney PA, Cook AJ, Miglioretti DL, et al. Use of clinical history affects accuracy of interpretive performance of screening mammography. J Clin Epidemiol 2012;65:219–230.
22. Waite S, Scott J, Gale B, et al. Interpretive error in radiology. AJR Am J Roentgenol 2017;208:739–749.
23. Jarvik JG, Deyo RA. Moderate versus mediocre: the reliability of spine MR data interpretations. Radiology 2009;250:15–17.
24. Wood KB, Khanna G, Vaccaro AR, et al. Assessment of two thoracolumbar fracture classification systems as used by multiple surgeons. J Bone Joint Surg Am 2005;87:1423–1429.
25. Arana E, Kovacs FM, Royuela A, et al. Agreement in the assessment of metastatic spine disease using scoring systems. Radiother Oncol 2015; 115:135–140.
26. Arana E, Kovacs FM, Royuela A, et al. Agreement in metastatic spinal cord compression. J Natl Compr Canc Netw 2016;14:70–76.
27. Fisher CG, Versteeg AL, Schouten R, et al. Reliability of the spinal instability neoplastic scale among radiologists: an assessment of instability secondary to spinal metastases. AJR Am J Roentgenol 2014;203:869–874.
28. Arana E, Kovacs FM, Royuela A, et al. Spine instability neoplastic score: agreement across different medical and surgical specialties. Spine J 2016; 16:591–599.
29. Khan L, Mitera G, Probyn L, et al. Inter-rater reliability between musculoskeletal radiologists and orthopedic surgeons on computed tomography imaging features of spinal metastases. Curr Oncol 2011;18:e282–287.

See JNCCN.org for supplemental online content.

JNCCN

National Comprehensive Cancer Network®

JOURNAL OF THE NATIONAL COMPREHENSIVE CANCER NETWORK

Supplemental online content for:

Metastatic Versus Osteoporotic Vertebral Fractures on MRI: A Blinded, Multicenter, and Multispecialty Observer Agreement Evaluation

Estanislao Arana, MD, MHE, PhD; Francisco M. Kovacs, MD, PhD; Ana Royuela, PhD; Beatriz Asenjo, MD, PhD; Fatima Nagib, MD; Sandra Pérez-Aguilera, MD; María Dejoz, BEng; Alberto Cabrera-Zubizarreta, MD; Yolanda García-Hidalgo, MD, PhD; Ana Estremera, MD, PhD; for the Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis

J Natl Compr Canc Netw 2020;18(3):267–273

eTable 1. Sequences for MRI Examinations

Pulse Sequence	TR/TE (ms)	FOV (mm)	MAX	NAV	Thickness (mm)	Comments	
Localizer	30/10	400	128 × 128	1	10	Flip angle 50°	Gradient echo
Sagittal T1	440–550/14–20	270	156–307 × 192–512	2	4	1.3–0.4 mm gap	Spin-echo
Sagittal T2	3,300–2,896/102.9–120	270	156–307 × 192–512	2	4	1.3–0.4 mm gap	Turbo spin-echo imaging, 12-echo train length
Sagittal STIR	3,000/45/150 (inversion time)	270	156–307 × 192–512	2	4–6	1.3–0.4 mm gap	Turbo spin-echo imaging, 12-echo train length
Axial T2	3,040–2,896/103–120	180	224–190 × 256–512	3	4	0.4 mm gap	Turbo spin-echo imaging, 5-echo train length

Abbreviations: FOV, field of view; MAX, matrix; NAV, number of signals acquired; STIR, short inversion time inversion-recovery; TE, echo time; TR, repetition time.

eTable 2. Imaging Findings Assessed

Imaging Finding	Possible Values
Pattern of signal abnormalities (pattern of replacement of normal vertebral signal with bone marrow edema)	"Partially or completely" vs "showing a bandlike pattern"
Horizontal fracture line on fluid-sensitive sequence (STIR) or T2-weighted images	"Yes" vs "no"
Deposit-like appearance of pedicle involvement	"Yes" vs "no"
Convexity of posterior vertebral body border (bulging posterior cortex)	"Yes" vs "no"
Posterosuperior retropulsion	"Yes" vs "no"
Symmetry of signal intensity changes	"Symmetrical" vs "asymmetrical"

Abbreviation: STIR, short inversion time inversion-recovery.

eTable 3. Interobserver Agreement on Imaging Findings

	Kappa (95% CI)
Radiology (n=9)	
Pattern of signal abnormalities	0.410 (0.351–0.473)
Horizontal fracture line	0.352 (0.277–0.432)
Deposit-like appearance of pedicle involvement	0.476 (0.422–0.534)
Bulging posterior cortex	0.602 (0.545–0.661)
Posterosuperior retropulsion	0.367 (0.312–0.424)
Symmetry of signal intensity changes	0.277 (0.229–0.327)
Neurosurgery (n=7)	
Pattern of signal abnormalities	0.428 (0.365–0.495)
Horizontal fracture line	0.130 (0.087–0.176)
Deposit-like appearance of pedicle involvement	0.473 (0.409–0.539)
Bulging posterior cortex	0.400 (0.339–0.464)
Posterosuperior retropulsion	0.445 (0.390–0.502)
Symmetry of signal intensity changes	0.267 (0.213–0.324)
Orthopedic surgery (n=5)	
Pattern of signal abnormalities	0.327 (0.270–0.386)
Horizontal fracture line	0.198 (0.145–0.253)
Deposit-like appearance of pedicle involvement	0.412 (0.340–0.487)
Bulging posterior cortex	0.104 (0.064–0.144)
Posterosuperior retropulsion	0.533 (0.467–0.602)
Symmetry of signal intensity changes	0.163 (0.109–0.219)
Radiation oncology (n=4)	
Pattern of signal abnormalities	0.355 (0.280–0.433)
Horizontal fracture line	0.326 (0.242–0.412)
Deposit-like appearance of pedicle involvement	0.416 (0.341–0.493)
Bulging posterior cortex	0.635 (0.561–0.711)
Posterosuperior retropulsion	0.101 (0.047–0.155)
Symmetry of signal intensity changes	0.388 (0.314–0.465)

eTable 4. Intraobserver Agreement on Imaging Findings

	Median Kappa (IQR)
Radiology (n=9)	
Pattern of signal abnormalities	0.722 (0.606–0.764)
Horizontal fracture line	0.639 (0.472–0.721)
Deposit-like appearance of pedicle involvement	0.707 (0.624–0.732)
Bulging posterior cortex	0.768 (0.640–0.800)
Posterosuperior retropulsion	0.673 (0.624–0.731)
Symmetry of signal intensity changes	0.575 (0.383–0.646)
Neurosurgery (n=7)	
Pattern of signal abnormalities	0.754 (0.533–0.894)
Horizontal fracture line	0.657 (0.458–0.914)
Deposit-like appearance of pedicle involvement	0.653 (0.527–0.914)
Bulging posterior cortex	0.844 (0.495–0.969)
Posterosuperior retropulsion	0.689 (0.617–0.941)
Symmetry of signal intensity changes	0.597 (0.402–0.902)
Orthopedic surgery (n=5)	
Pattern of signal abnormalities	0.549 (0.510–0.555)
Horizontal fracture line	0.457 (0.399–0.515)
Deposit-like appearance of pedicle involvement	0.504 (0.460–0.549)
Bulging posterior cortex	0.682 (0.618–0.693)
Posterosuperior retropulsion	0.712 (0.587–0.719)
Symmetry of signal intensity changes	0.409 (0.360–0.460)
Radiation oncology (n=4)	
Pattern of signal abnormalities	0.646 (0.603–0.808)
Horizontal fracture line	0.486 (0.433– 0.712)
Deposit-like appearance of pedicle involvement	0.666 (0.592–0.797)
Bulging posterior cortex	0.724 (0.639–0.846)
Posterosuperior retropulsion	0.586 (0.429–0.775)
Symmetry of signal intensity changes	0.584 (0.517–0.753)

Abbreviation: IQR, interquartile range.

eTable 5. Diagnostic Accuracy[a]

	N	Median Kappa (IQR)
All readers		
Cancer history undisclosed	25	0.437 (0.326–0.511)
Cancer history disclosed	25	0.443 (0.398–0.526)
Specialty		
Neurosurgery		
Cancer history undisclosed	7	0.327 (0.230–0.511)
Cancer history disclosed	7	0.411 (0.314–0.534)
Radiation oncology		
Cancer history undisclosed	4	0.446 (0.348–0.507)
Cancer history disclosed	4	0.435 (0.354–0.490)
Orthopedic surgery		
Cancer history undisclosed	5	0.368 (0.325–0.445)
Cancer history disclosed	5	0.398 (0.311–0.444)
Radiology		
Cancer history undisclosed	9	0.437 (0.414–0.525)
Cancer history disclosed	9	0.484 (0.443–0.526)
Hospital category (complexity)[b]		
Category 2		
Cancer history undisclosed	2	0.381 (0.325–0.437)
Cancer history disclosed	2	0.372 (0.311–0.433)
Category 3		
Cancer history undisclosed	9	0.470 (0.403–0.525)
Cancer history disclosed	9	0.484 (0.410–0.534)
Category 4		
Cancer history undisclosed	7	0.445 (0.327–0.565)
Cancer history disclosed	7	0.437 (0.411–0.543)
Category 5		
Cancer history undisclosed	7	0.413 (0.281–0.426)
Cancer history disclosed	7	0.443 (0.359–0.526)
Years of experience		
≤7		
Cancer history undisclosed	7	0.403 (0.325–0.437)
Cancer history disclosed	7	0.411 (0.359–0.491)
8–13		
Cancer history undisclosed	6	0.397 (0.253–0.445)
Cancer history disclosed	6	0.421 (0.314–0.526)
≥14		
Cancer history undisclosed	12	0.491 (0.428–0.543)
Cancer history disclosed	12	0.477 (0.435–0.554)

Abbreviations: IQR, interquartile range; MVF, metastatic vertebral fracture; OVF, osteoporotic vertebral fracture.
[a]Defined as the concordance between each reader's diagnosis at the first round (OVF vs MVF) and the reference diagnosis (established through biopsy or follow-up >6 months).
[b]Based on size, availability of high-tech medical equipment and procedures, and degree of educational activity. No readers from category 1 hospitals (simplest) were included in this study.

eTable 6. Diagnostic Accuracy[a] Depending on Presence of Preexisting Fractures and Disclosure of Clinical History

	Median Kappa (IQR)
Cases without preexisting fractures	
Before clinical history of cancer was disclosed	0.452 (0.387–0.509)
After clinical history of cancer was disclosed	0.462 (0.407–0.570)
Cases with preexisting fractures	
Before clinical history of cancer was disclosed	0.286 (0.183–0.396)
After clinical history of cancer was disclosed	0.331 (0.219–0.368)

Abbreviations: IQR, interquartile range; MVF, metastatic vertebral fracture; OVF, osteoporotic vertebral fracture.
[a]Diagnostic accuracy is defined as the concordance between each reader's diagnosis at the first round (OVF vs MVF) and the reference diagnosis (established through biopsy or follow-up >6 months).

eAppendix 1.

Members of the Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis (in alphabetical order)

Ana Alonso[1,2]; Marco Antonio Álvarez[1,3]; Luis Álvarez-Galovich[1,4]; Aida Antuña[1,3]; Joaquín Cabrera[1,5]; Carlos Casillas[1,6]; Gregorio Catalán[7,8]; Diego Dualde[7,9]; Nicomedes Fernández-Baillo[7,10]; Antonio Ferreiro[7,11]; Pilar Ferrer[1,12]; Sara García-Duque[7,13]; Cristina García-Villar[7,14]; Ovidio Hernando-Requejo[1,15]; Laín Ibáñez[1,16]; Ana Lersundi[1,17]; Marta Manero[1,18]; Antonio Martín[1,19]; Julio César Palomino[7,20]; Luis A. Pérez-Romasanta[1,21]; Julio Plata-Bello[1,22]; Raquel Prada[1,20]; Héctor Roldán[1,22]; Luis Maria Romero-Muñoz[1,23]; Félix Tomé-Bermejo[1,4]; Vicente Vanaclocha[1,24]; and Joaquín Zamarro[7,25]

[1]Spanish Back Pain Research Network, Kovacs Foundation, Palma de Mallorca, Spain
[2]Hospital Universitario Rey Juan Carlos, Móstoles, Madrid, Spain
[3]Hospital Universitario Central de Asturias, Asturias, Spain
[4]Hospital Universitario Fundación Jiménez Díaz, Madrid, Spain
[5]Hospital Universitario de Badajoz, Badajoz, Spain
[6]Hospital Jaume I, Castellón, Spain
[7]Unidad de la Espalda Kovacs, Hospital Universitario HLA-Moncloa, Madrid
[8]Hospital de Cruces, Baracaldo, Spain
[9]Hospital Clínico Universitario de Valencia, Valencia, Spain
[10]Hospital La Paz, Madrid, Spain
[11]Hospital de Madrid, HM Hospitales, Madrid, Spain
[12]Hospital Intermutual de Levante, San Antonio de Benagéber, Valencia, Spain
[13]Hospital Universitario HM Sanchinarro, Madrid, Spain
[14]Hospital Universitario Puerta del Mar, Cádiz, Spain
[15]Hospital Universitario HM Puerta del Sur, Móstoles, Madrid, Spain
[16]Hospital Universitario 12 de Octubre, Madrid, Spain
[17]Hospital Universitario Donostia, Donostia, Gipuzkoa, Spain
[18]Clínica Vistahermosa, Alicante, Spain
[19]Hospital Doctor Peset, Valencia, Spain
[20]Hospital POVISA, Vigo, Spain
[21]Hospital Universitario de Salamanca, Salamanca, Spain
[22]Hospital Universitario de Canarias, Santa Cruz de Tenerife, Spain
[23]Hospital Nacional de Parapléjicos, Toledo, Spain
[24]Hospital General Universitario de Valencia, Valencia, Spain
[25]Hospital Universitario Virgen de la Arrixaca, Murcia, Spain

Anexo

Aceptación por escrito y con firma original de los coautores para que el doctorando presente el trabajo

D. Francisco Kovacs, coautor de los trabajos:

- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J; SpanishBack Pain Research Network Task Force for the Improvement of Inter-DisciplinaryManagement of Spinal Metastasis. Agreement in Metastatic Spinal Cord Compression.J Natl ComprCancNetw. 2016 Jan;14(1):70-6. Epub 2016 Jan 5. PubMed PMID:26733556.
- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J; SpanishBack Pain Research Network Task Force for the Improvement of Inter-DisciplinaryManagement of Spinal Metastasis. Spine Instability Neoplastic Score: agreementacross different medical and surgical specialties. Spine J. 2016 May;16(5):591-9.doi: 10.1016/j.spinee.2015.10.006. Epub 2015 Oct 22. PubMed PMID: 26471708.
- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez U, Zamora J; SpanishBack Pain Research Network Task Force for the improvement of inter-disciplinarymanagement of spinal metastasis. Agreement in the assessment of metastatic spine disease using scoring systems. Radiother Oncol. 2015 Apr;115(1):135-40. doi:10.1016/j.radonc.2015.03.016. Epub 2015 Apr 10. PubMed PMID: 25869337.
- Arana E, Kovacs FM, Royuela A, Asenjo B,Nagib F, Pérez-Aguilera S, Dejoz M, Cabrera-Zubizarreta A, García-Hidalgo Y, Estremera A. Metastatic versus osteoporotic vertebral fractures on MRI: a blinded, multicenter and multispecialty observer agreement evaluation J Natl Compr Canc Netw. 2020, aceptado, en prensa.

Acepto que Estanislao Arana, doctorando por la Universidad de Salamanca con la tesis " Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen ", presente el trabajo, y declaro que el doctorando es el autor principal de la investigación recogida en los artículos.

Asimismo, presento renuncia expresa a presentar los trabajos como parte de otra Tesis Doctoral.

Firmado

En Madrid, a diecisiete de diciembre de 2019

Dª Beatriz Asenjo García coautor de los trabajos:

- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J; Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis. Agreement in Metastatic Spinal Cord Compression. J Natl Compr Canc Netw. 2016 Jan;14(1):70-6. Epub 2016 Jan 5. PubMed PMID:26733556.
- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J; Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis. Spine Instability Neoplastic Score: agreement across different medical and surgical specialties. Spine J. 2016 May;16(5):591-9.doi: 10.1016/j.spinee.2015.10.006. Epub 2015 Oct 22. PubMed PMID: 26471708.
- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez U, Zamora J; Spanish Back Pain Research Network Task Force for the improvement of inter-disciplinary management of spinal metastasis. Agreement in the assessment of metastatic spine disease using scoring systems. Radiother Oncol. 2015 Apr;115(1):135-40. doi:10.1016/j.radonc.2015.03.016. Epub 2015 Apr 10. PubMed PMID: 25869337.

- Arana E, Kovacs FM, Royuela A, Asenjo B, Nagib F, Pérez-Aguilera S, Dejoz M, Cabrera-Zubizarreta A, García-Hidalgo Y, Estremera A. Metastatic versus osteoporotic vertebral fractures on MRI: a blinded, multicenter and multispecialty observer agreement evaluation J Natl Compr Canc Netw. 2020, aceptado, en prensa.

Acepto que Estanislao Arana, doctorando por la Universidad de Salamanca con la tesis " Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen ", presente el trabajo y, declaro que el doctorando es el autor principal de la investigación recogida en los artículos.

Asimismo, presento renuncia expresa a presentar los trabajos como parte de otra Tesis Doctoral.

Firmado

En Málaga, a 23 diciembre de 2019

D/Dª coautor de los trabajos:

- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J; Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis. Agreement in Metastatic Spinal Cord Compression. J Natl Compr Canc Netw. 2016 Jan;14(1):70-6. Epub 2016 Jan 5. PubMed PMID:26733556.
- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J; Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis. Spine Instability Neoplastic Score: agreement across different medical and surgical specialties. Spine J. 2016 May;16(5):591-9.doi: 10.1016/j.spinee.2015.10.006. Epub 2015 Oct 22. PubMed PMID: 26471708.
- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramirez U, Zamora J; Spanish Back Pain Research Network Task Force for the improvement of inter-disciplinary management of spinal metastasis. Agreement in the assessment of metastatic spine disease using scoring systems. Radiother Oncol. 2015 Apr;115(1):135-40. doi:10.1016/j.radonc.2015.03.016. Epub 2015 Apr 10. PubMed PMID: 25869337.
- Arana E, Kovacs FM, Royuela A, Asenjo B, Nagib F, Pérez-Aguilera S, Dejoz M, Cabrera-Zubizarreta A, García-Hidalgo Y, Estremera A. Metastatic versus osteoporotic vertebral fractures on MRI: a blinded, multicenter and multispecialty observer agreement evaluation J Natl Compr Canc Netw. 2020, aceptado, en prensa.

Acepto que Estanislao Arana, doctorando por la Universidad de Salamanca con la tesis " Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen ", presente el trabajo y, declaro que el doctorando es el autor principal de la investigación recogida en los artículos.

Asimismo, presento renuncia expresa a presentar los trabajos como parte de otra Tesis Doctoral.

Firmado Ana Royuela

En 13 ,a diciembre de 2019

D/Dª coautor de los trabajos: Úrsula Pérez Ramírez

- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramirez Ú, Zamora J; Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis. Agreement in Metastatic Spinal Cord Compression. J Natl Compr Canc Netw. 2016 Jan;14(1):70-6. Epub 2016 Jan 5. PubMed PMID:26733556.
- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J; Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis. Spine Instability Neoplastic Score: agreement across different medical and surgical specialties. Spine J. 2016 May;16(5):591-9.doi: 10.1016/j.spinee.2015.10.006. Epub 2015 Oct 22. PubMed PMID: 26471708.
- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramirez U, Zamora J; Spanish Back Pain Research Network Task Force for the improvement of inter-disciplinary management of spinal metastasis. Agreement in the assessment of metastatic spine disease using scoring systems. Radiother Oncol. 2015 Apr;115(1):135-40. doi:10.1016/j.radonc.2015.03.016. Epub 2015 Apr 10. PubMed PMID: 25869337.

Acepto que Estanislao Arana, doctorando por la Universidad de Salamanca con la tesis " Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen ", presente el trabajo y, declaro que el doctorando es el autor principal de la investigación recogida en los artículos.

Asimismo, presento renuncia expresa a presentar los trabajos como parte de otra Tesis Doctoral.

Firmado

Úrsula PR.

En Valencia, a 14 de diciembre de 2019

D/Dª coautor de los trabajos:

- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J; Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis. Agreement in Metastatic Spinal Cord Compression. J Natl Compr Canc Netw. 2016 Jan;14(1):70-6. Epub 2016 Jan 5. PubMed PMID:26733556.
- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J; Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis. Spine Instability Neoplastic Score: agreement across different medical and surgical specialties. Spine J. 2016 May;16(5):591-9.doi: 10.1016/j.spinee.2015.10.006. Epub 2015 Oct 22. PubMed PMID: 26471708.
- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez U, Zamora J; Spanish Back Pain Research Network Task Force for the improvement of inter-disciplinary management of spinal metastasis. Agreement in the assessment of metastatic spine disease using scoring systems. Radiother Oncol. 2015 Apr;115(1):135-40. doi:10.1016/j.radonc.2015.03.016. Epub 2015 Apr 10. PubMed PMID: 25869337.
- Arana E, Kovacs FM, Royuela A, Asenjo B, Nagib F, Pérez-Aguilera S, Dejoz M, Cabrera-Zubizarreta A, Garcia-Hidalgo Y, Estremera A. Metastatic versus osteoporotic vertebral fractures on MRI: a blinded, multicenter and multispecialty observer agreement evaluation J Natl Compr Canc Netw. 2020, aceptado, en prensa.

Acepto que Estanislao Arana, doctorando por la Universidad de Salamanca con la tesis " Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen ", presente el trabajo y, declaro que el doctorando es el autor principal de la investigación recogida en los artículos.

Asimismo, presento renuncia expresa a presentar los trabajos como parte de otra Tesis Doctoral.

Firmado Yolanda García Hidalgo

En 16 ,a diciembre de 2019
Madrid

D/Dª coautor de trabajo: Sandra Pérez-Aguilera

- Arana E, Kovacs FM, Royuela A, Asenjo B, Nagib F, Pérez-Aguilera S, Dejoz M, Cabrera-Zubizarreta A, García-Hidalgo Y, Estremera A. Metastatic versus osteoporotic vertebral fractures on MRI: a blinded, multicenter and multispecialty observer agreement evaluation J Natl Compr Canc Netw. 2020, aceptado, en prensa.

Acepto que Estanislao Arana, doctorando por la Universidad de Salamanca con la tesis " Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen ", presente el trabajo y, declaro que el doctorando es el autor principal de la investigación recogida en los artículos.

Asimismo, presento renuncia expresa a presentar los trabajos como parte de otra Tesis Doctoral.

Firmado

En Manacor ,a 23 diciembre de 2019

D/Dª Fátima Nagib coautor de trabajo:

- Arana E, Kovacs FM, Royuela A, Asenjo B, Nagib F, Pérez-Aguilera S, Dejoz M, Cabrera-Zubizarreta A, García-Hidalgo Y, Estremera A. Metastatic versus osteoporotic vertebral fractures on MRI: a blinded, multicenter and multispecialty observer agreement evaluation J Natl Compr Canc Netw. 2020, aceptado, en prensa.

Acepto que Estanislao Arana, doctorando por la Universidad de Salamanca con la tesis " Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen ", presente el trabajo y, declaro que el doctorando es el autor principal de la investigación recogida en los artículos.

Asimismo, presento renuncia expresa a presentar los trabajos como parte de otra Tesis Doctoral.

Firmado

En Málaga ,a 30 diciembre de 2019

Ana Estremera Rodrigo

D/Dª coautor de trabajo:

- Arana E, Kovacs FM, Royuela A, Asenjo B, Nagib F, Pérez-Aguilera S, Dejoz M, Cabrera-Zubizarreta A, García-Hidalgo Y, Estremera A. Metastatic versus osteoporotic vertebral fractures on MRI: a blinded, multicenter and multispecialty observer agreement evaluation J Natl Compr Canc Netw. 2020, aceptado, en prensa.

Acepto que Estanislao Arana, doctorando por la Universidad de Salamanca con la tesis " Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen ", presente el trabajo y, declaro que el doctorando es el autor principal de la investigación recogida en los artículos.

Asimismo, presento renuncia expresa a presentar los trabajos como parte de otra Tesis Doctoral.

Firmado

En Palma, a 16 diciembre de 2019

D/Dª Mª Cristina Dejoz coautor de trabajo:

- Arana E, Kovacs FM, Royuela A, Asenjo B, Nagib F, Pérez-Aguilera S, Dejoz M, Cabrera-Zubizarreta A, García-Hidalgo Y, Estremera A. Metastatic versus osteoporotic vertebral fractures on MRI: a blinded, multicenter and multispecialty observer agreement evaluation J Natl Compr Canc Netw. 2020, aceptado, en prensa.

Acepto que Estanislao Arana, doctorando por la Universidad de Salamanca con la tesis " Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen ", presente el trabajo y, declaro que el doctorando es el autor principal de la investigación recogida en los artículos.

Asimismo, presento renuncia expresa a presentar los trabajos como parte de otra Tesis Doctoral.

Firmado

En Valencia ,a 26 diciembre de 2019

D. Alberto Cabrera Zubizarreta coautor de trabajo:

- Arana E, Kovacs FM, Royuela A, Asenjo B, Nagib F, Pérez-Aguilera S, Dejoz M, Cabrera-Zubizarreta A, Garcia-Hidalgo Y, Estremera A. Metastatic versus osteoporotic vertebral fractures on MRI: a blinded, multicenter and multispecialty observer agreement evaluation J Natl Compr Canc Netw. 2020, aceptado, en prensa.

Acepto que Estanislao Arana, doctorando por la Universidad de Salamanca con la tesis " Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen ", presente el trabajo y, declaro que el doctorando es el autor principal de la investigación recogida en los artículos.

Asimismo, presento renuncia expresa a presentar los trabajos como parte de otra Tesis Doctoral.

Firmado

En Getxo, a 21 diciembre de 2019

D/Dª JAVIER ZAMORA ROMERO, coautor de los trabajos:

- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J; Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis. Agreement in Metastatic Spinal Cord Compression. J Natl Compr Canc Netw. 2016 Jan;14(1):70-6. Epub 2016 Jan 5. PubMed PMID:26733556.
- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez Ú, Zamora J; Spanish Back Pain Research Network Task Force for the Improvement of Inter-Disciplinary Management of Spinal Metastasis. Spine Instability Neoplastic Score: agreement across different medical and surgical specialties. Spine J. 2016 May;16(5):591-9.doi: 10.1016/j.spinee.2015.10.006. Epub 2015 Oct 22. PubMed PMID: 26471708.
- Arana E, Kovacs FM, Royuela A, Asenjo B, Pérez-Ramírez U, Zamora J; Spanish Back Pain Research Network Task Force for the improvement of inter-disciplinary management of spinal metastasis. Agreement in the assessment of metastatic spine disease using scoring systems. Radiother Oncol. 2015 Apr;115(1):135-40. doi:10.1016/j.radonc.2015.03.016. Epub 2015 Apr 10. PubMed PMID: 25869337.

Acepto que Estanislao Arana, doctorando por la Universidad de Salamanca con la tesis " Columna metastásica: diagnóstico y acuerdo interobservador en diagnóstico por la imagen ", presente el trabajo y, declaro que el doctorando es el autor principal de la investigación recogida en los artículos.

Asimismo, presento renuncia expresa a presentar los trabajos como parte de otra Tesis Doctoral.

En Madrid a 5 de diciembre de 2019

Fdo. Javier Zamora Romero

ZAMORA ROMERO JAVIER - 10071879H

Firmado digitalmente por ZAMORA ROMERO JAVIER - 10071879H
Fecha: 2019.12.05 10:21:47 +01'00'