The Spine Journal 16 (2016) 591–599

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
1529-9430/© 2015 Elsevier Inc. All rights reserved.

CONCLUSIONS: Agreement on the assessment of metastatic spine instability is moderate. The SINS can help improve communication among clinicians in oncology care. © 2015 Elsevier Inc. All rights reserved.

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.

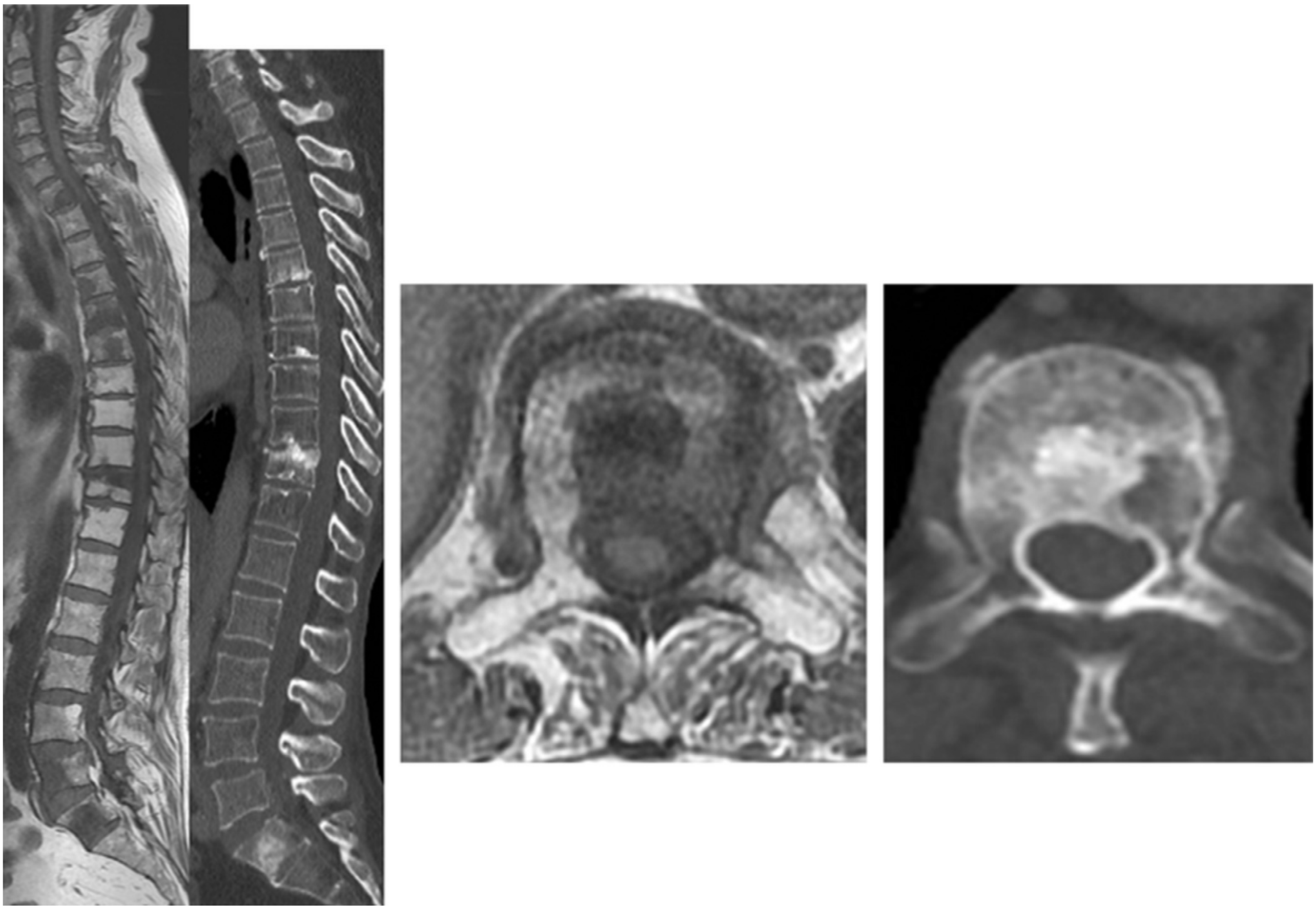

Figure 1. An example of the information pack provided to readers for each patient.
Images corresponding to a 69-year-old woman suffering from breast cancer, who reported continuous back pain without referred pain. She presented lung, liver, and bone metastases. Please select the most unstable spine level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