Review Article

Intra- and interobserver reliability of the Spinal Instability Neoplastic Score system for instability in spine metastases: a systematic review and meta-analysis

Zach Pennington, A. Karim Ahmed, Ethan Cottrill, Erick M. Westbroek, Matthew L. Goodwin, Daniel M. Sciubba

Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Contributions: (I) Conception and design: Z Pennington, DM Sciubba, ML Goodwin; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: Z Pennington, E Cottrill, AK Ahmed; (V) Data Analysis and interpretation: Z Pennington, EM Westbroek; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Daniel M. Sciubba, MD. 600 N. Wolfe Street, Meyer 5-185A, Baltimore, MD 21287, USA. Email: dsciubb1@jhmi.edu.

Abstract: Mechanical instability is one of the two main indications for surgical intervention in patients with metastatic spine disease. Since its publication in 2010, the Spinal Instability Neoplastic Score (SINS) has been the most commonly used means of assessing mechanical instability. To prove clinically valuable though, diagnostic tests must demonstrate consistency across measures and across observers. Here we report a systematic review and meta-analysis of all prior reports of intraobserver and interobserver reliability of the SINS score. To identify articles, we queried the PubMed, CINAHL, EMBASE, Cochrane, and Web of Science databases for all full-text English articles reporting interobserver or intraobserver reliability for the SINS score, category, or a domain of the SINS score. Articles reporting confidence intervals for these metrics were then subjected to meta-analysis to identify pooled estimates of reliability. Of 167 unique studies identified, seven met inclusion criteria and were subjected to qualitative review and meta-analysis. Intraobserver reliability for SINS score was found to be near perfect [estimate =0.815; 90% CI (0.661–0.969)] and interobserver reliability was substantial [0.673; (0.227–1.12)]. Intraobserver and interobserver reliability among spine surgeons was significantly better than reliability across all observers (both P<0.0001). Qualitative analysis suggested that increased surgeon experience may be associated with greater intraobserver and interobserver reliability among spine surgeons. On the whole, meta-analysis of the available literature suggests SINS to have good intraobserver and interobserver reliability, giving it the potential to be a valuable guide to the management of patients with spinal metastases. Further research is required to demonstrate that SINS score correlates with the clinical decision to stabilize.

Keywords: Evidence-based medicine; osteolysis; mechanical instability; Spinal Instability Neoplastic Score (SINS); spine metastasis

Submitted Jan 17, 2019. Accepted for publication Jan 24, 2019.
doi: 10.21037/atm.2019.01.65
View this article at: http://dx.doi.org/10.21037/atm.2019.01.65

Introduction

The skeleton is the third most common site of metastasis in cancer (1) with upwards of 70% of patients having pathological evidence of metastatic disease to the spine at the time of death (2,3). As these metastases represent disseminated disease, treatment designed to specifically address these metastatic sites is typically palliative rather than curative in nature. For patients with metastatic disease of the spine the two major presenting symptoms requiring intervention are pain and neurological dysfunction (4). The former can be oncologic in nature—pain attributable to biochemical alterations of the bony microenvironment—

Table 1 Inclusion and exclusion criteria

Inclusion
Full text in English
Primary literature—format is cohort, randomized controlled trial, case series
Reports validation/reliability of Spinal Instability Neoplastic Score, i.e., intra-observer or interobserver reliability
Minimum of five observers used for evaluation of interobserver reliability
Observers all given the same images for scoring
All images are from patients with biopsy-confirmed metastatic spine disease
Exclusion
Abstract only/conference proceeding only
Article format is review, perspective, commentary, case report (<5 patients)
Article does not report quantitative results of interobserver or intraobserver reliability testing

or mechanical in nature—pain that is worsened by movement (5). Mechanical pain is particularly concerning to the spine surgeon and prompts immediate assessment of the mechanical stability of the affected vertebral segment.

Formally, mechanical instability of the spine is defined as a "non-optimal state of equilibrium" (6) or in the context of metastatic disease, a "loss of spinal integrity as a result of a neoplastic process that is associated with movement-related pain, symptomatic or progressive deformity and/or neural compromise under physiological loads" (7). In practice this describes a vertebral body that will progressively fracture if not stabilized. Despite this formalized definition which has existed in print for more than three decades, until the late 2000s, there existed no standard means of assessing spinal column instability. Then in 2010, the Spine Oncology Study Group published the Spinal Instability Neoplastic Score (SINS), a formalized scoring system designed to allow for the uniform assessment of mechanical instability in the context of metastatic spine disease (7). The system scores lesions on a scale from 0–18 using six variables—pain, location, bone lesion quality (lytic/blastic), alignment, vertebral body collapse, and posterolateral element involvement. Lesions are then described as stable [0–6], potentially unstable [7–12], or unstable [13–18]. As with any scoring system, the utility of SINS is determined by its ability to accurately guide practice and to yield consistent results both across and within reviewers. Here we systematically review the medical literature for studies describing the intraobserver and interobserver reliability of the SINS score and perform a meta-analysis of reliability scores across all observers.

Methods

We queried the medical literature for all reports describing the intraobserver and/or interobserver reliability of the Spinal Instability Neoplastic Score published as of November 5th, 2018. Inclusion and exclusion criteria are available as *Table 1*. All included studies were sourced from the English primary literature with full-text availability and included a minimum of five independent reviewers when evaluating interobserver reliability. Studies were excluded if they did not meet the inclusion criteria, reported qualitative results only, did not have full English text availability, or were non-primary literature (e.g., reviews, perspectives, commentaries, case reports). Databases included in our search were: PubMed, CINAHL, EMBASE, Cochrane, and Web of Science. All articles were screened by two reviewers (ZP and EC) and in cases of conflict, a third reviewer (AKA) was involved to resolve the conflict. Studies included in the full text review were then evaluated for the presence of one of the endpoints of interest, namely a quantitative measure of intraobserver or interobserver reliability for the SINS scoring system.

Meta-analysis

A meta-analysis was conducted for intra-observer and inter-observer reliability for overall SINS score, SINS categorization (stable, potentially unstable, or unstable), and each SINS domain (pain, location, bone lesion quality, alignment, collapse, and posterolateral involvement). Studies were included if they provided confidence

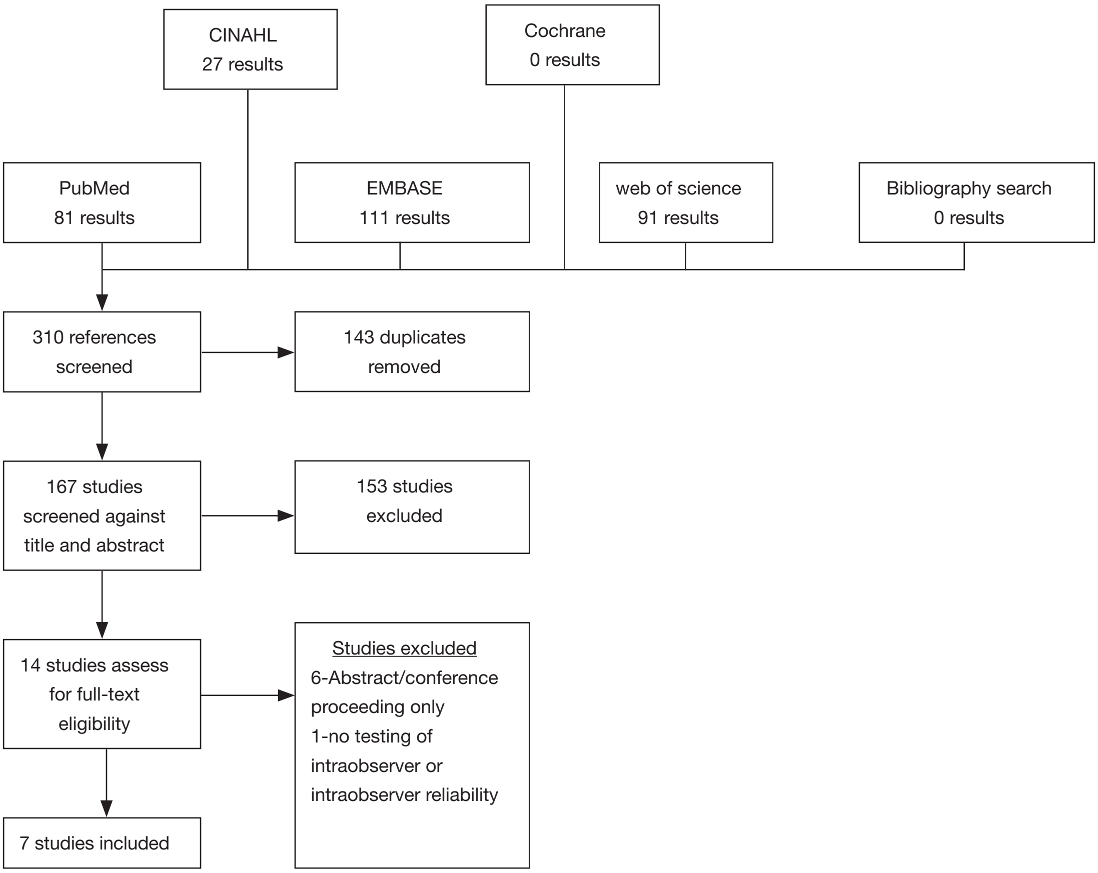

Figure 1 PRISMA diagram for results of literature review.

intervals in addition to a point estimate. In addition to evaluating reliability across all observers, we also pooled results for specific specialties (e.g., spine surgeons) where possible. Final estimates of reliability were calculated using Microsoft Excel® (Redmond, WA) by calculating the individual variances for each point estimate included based upon the confidence interval limits and percentage (e.g., 95% CI). Mean estimate for the statistic was calculated as the weighted mean of the statistics reported for included studies and pooled variance was calculated as the sum of the variance of the included statistics and the mean of the calculated variances for those statistics as described by Rudmin (8). When pooling results, study statistics were weighted by the number of observers. In cases where studies shared observers [i.e., (9,10)], duplicated observers were included in the first study only. If duplicated observers could not be excluded, then only the larger of the two studies sharing the observers was included in the meta-analysis. Pooled estimates are categorized according to the method of Landis and Cook as "almost perfect" (0.81–1.00), "substantial" (0.61–0.80), "moderate" (0.41–0.60), "fair" (0.21–0.40), and "slight" (0.00–0.20) (11). To compare reliability statistics between groups we employed independent t-tests. Alpha level was set at 0.05 *a priori*.

Results

The results of our query are illustrated in *Figure 1* as a PRISMA diagram. We identified 167 unique studies, of which 14 were eligible for full-text review. Of these 14 articles, 7 were excluded—6 studies were excluded for being abstracts/conference presentations only (12-17) and 1 study was excluded because it did not report any quantitative results of intraobserver or interobserver reliability (18). The seven included studies reported the results for 236 unique reviewers evaluating 250 patient images (9,10,19-23). Of studies reporting demographic information, mean patient age was 60.6 years and 51.7% of the cohort was male (9,10,19). Six studies reported the location of the evaluated

Table 2 Summary of evidence

Study	Methods
Arana *et al.*, 2016 (19)	90 patients with biopsy-confirmed spinal metastasis; 83 observers from 61 departments; 30 spine surgeons
Campos *et al.*, 2014 (20)	30 patients with biopsy-confirmed spinal metastases; 6 observers from 3 departments; 3 spine surgeons
Fisher *et al.*, 2014a (10)	30 patients; 33 radiation oncologists across 10 sites; 11 spine surgeons with spinal oncology training used as gold standard
Fisher *et al.*, 2014b (9)	30 patients (same as Fisher *et al.*, 2014a cohort); 37 radiologists (13 with neuroradiology fellowship training); 11 spine surgeons with spinal oncology training used as gold standard
Fourney *et al.*, 2011 (21)	30 patients; 24 SOSG members
Fox *et al.*, 2017 (22)	30 patients; 23 ortho or neurosurgery residents and 2 spine fellows
Teixeira *et al.*, 2013 (23)	40 cases; 17 physicians; 10 spine surgeons

metastasis with 17% being cervical, 54% being thoracic, and 29% being lumbar (9,10,19-22). Four studies reported the primary lesion pathology with the most common lesion primaries being breast (32%), lung (16.7%), and prostate (14.7%) (9,10,19,20).

The results of all included studies are available in *Table 2* and *Table 3*. Of the four studies reporting both intraobserver and interobserver agreement for overall SINS score, intraobserver agreement varied between 0.767 and 0.96 and interobserver agreement varied between 0.546 and 0.99 (19-22). A fifth study (23) reported the results of interobserver testing only (κ=0.375) and found it to be significantly lower than intraobserver and interobserver reliability reported in the other four studies. Intraobserver and interobserver reliability for SINS category were found to be slightly worse, with all three studies reporting this outcome demonstrating only "substantial" agreement between reviewers for intraobserver agreement (0.605–0.68) and demonstrating "moderate" agreement for interobserver agreement (0.424–0.54) (9,10,19). Five studies reported the results of both intraobserver and interobserver agreement testing for the individual SINS domains (9,10,20-22) and a sixth study reported the results of interobserver agreement alone for the SINS domain (23). Among these studies, overall intraobserver and interobserver agreement varied widely dependent upon the category considered. Intraobserver agreement was near perfect for location (0.806–0.98) and pain (0.814–0.98), moderate to near perfect for bone quality (0.576–0.87), vertebral body collapse (0.590–0.92) and posterolateral element involvement (0.58–0.86), and substantial to near perfect for alignment (0.610–0.88). Interobserver agreement was generally lower across all domains, being substantial to near perfect for location (0.719–0.948), fair to substantial for bone quality (0.210–0.65), moderate to near perfect for pain (0.419–0.88), moderate for alignment (0.42–0.553), moderate to substantial for collapse (0.421–0.61), and fair to moderate for posterolateral involvement (0.295–0.55).

Six studies reported sub-analysis of intraobserver agreement by physician specialty or training level for overall SINS score or category. Fourney *et al.* and Fox *et al.* both reported near perfect intraobserver (0.886–0.907) and interobserver (0.846–0.99) agreement for overall SINS score among spine surgeons (21,22). Both Arana *et al.* and Campos *et al.* reported interobserver reliability (0.629–0.860) among orthopaedic surgeons to be substantial to near perfect (19,20). Arana *et al.* and Fisher *et al.* reported intraobserver and interobserver agreement for radiation oncologists, finding intraobserver agreement to be moderate to substantial (0.578–0.65) and interobserver agreement to be moderate (0.462–0.54) (10,19). Lastly, Arana *et al.* and Fisher *et al.* reported the results among radiologists, finding intraobserver reliability to be substantial (0.646–0.69) though interobserver reliability was only fair to moderate (0.328–0.53) (9,19). Only two studies reported differences in agreement by level of training (19,23). Arana *et al.* found no significant difference in intraobserver or interobserver agreement for overall SINS score—0.732–0.799 and 0.511–0.565, respectively—or SINS category—0.594–0.633 and 0.345–0.511, respectively—as a function of years of practice (19). To contrast this, Teixeira and colleagues did observe a difference, noting higher overall agreement for SINS score (κ=0.631 *vs.* 0.329) among experienced compared to inexperienced spine surgeons. This was driven

Table 3 Intraobserver and interobserver reliability (κ) estimates from included studies

Study	No. of Obs	Intraobserver agreement								No. of Obs	Interobserver agreement							
		Score	Cat	Loc	Pain	BQ	Aln	Clp	PLI		Score	Cat	Loc	Pain	BQ	Aln	Clp	PLI
Arana *et al.*, 2016 (19)	83	0.767	0.605	–	–	–	–	–	–	83	0.546	0.424	–	–	–	–	–	–
Campos *et al.*, 2014 (20)	30	0.96	–	0.98	0.98	0.87	0.88	0.92	0.86	30	0.79	–	0.811	0.587	0.210	0.453	0.421	0.295
Fisher *et al.*, 2014a (10)	44	–	0.65	0.96	0.91	0.68	0.63	0.63	0.58	33	–	0.54	0.94	0.88	0.55	0.42	0.57	0.43
Fisher *et al.*, 2014b (9)	48	–	0.68	0.96	0.85	0.75	0.69	0.71	0.68	37	–	0.53	0.94	0.73	0.65	0.49	0.61	0.55
Fourney *et al.*, 2011 (21)	24	0.886	–	0.806	0.859	0.528	0.614	0.590	0.662	24	0.846	–	0.790	0.841	0.244	0.456	0.462	0.492
Fox *et al.*, 2017 (22)	25	0.907	–	0.954	0.814	0.576	0.610	0.671	0.561	25	0.99	–	0.948	0.739	0.382	0.427	0.550	0.435
Teixeira *et al.*, 2013 (23)	17	–	–	–	–	–	–	–	–	17	0.375	–	0.719	0.419	0.220	0.553	0.428	0.424

Aln, alignment subscore; BQ, bone quality subscore; Cat, SINS Category; Clp, collapse subscore; Loc, location subscore; Obs, observers; Score, SINS Score; Pain, pain subscore; PLI, posterolateral involvement subscore.

mainly by greater overall agreement for spine location (κ=0.908 *vs.* 0.769), vertebral body involvement (κ=0.578 *vs.* 0.425), and posterior element involvement (κ=0.571 *vs.* 0.422) (23). They did not conduct inferential statistics to compare these groups though.

Meta-analysis

All seven studies met criteria to be included in the meta-analysis for at least one endpoint (*Table 4*). Overall, intraobserver reliability for SINS score was found to be near perfect (estimate = 0.815; 90% CI, 0.661–0.969) and interobserver reliability was substantial (0.673; 95% CI, 0.227–1.12). Agreement for SINS category was slightly worse both within (0.636; 95% CI, 0.484–0.789) and between observers (0.475; 95% CI, 0.188–0.762). Among the SINS domains, intraobserver agreement was best for location (0.882; 95% CI, 0.756–1.01) and pain (0.836; 95% CI, 0.782–0.889) and worst for bone lesion quality (0.552; 95% CI, 0.477–0.628). Interobserver agreement was also greatest for location (0.831; 95% CI, 0.643–1.02) and pain (0.694; 95% CI, 0.340–1.05) and poorest for bone quality (0.290; 95% CI, 0.147–0.433). Sub-analysis for agreement by specialty demonstrated spine surgeons (orthopaedic and/or neurosurgical specialization) to have significantly higher interobserver reliability as compared to both orthopaedic surgeons without specific spine specialization for overall SINS score (0.919 *vs.* 0.625; P<0.0001) and the overall cohort (0.919 *vs.* 0.673; P<0.0001). Spine surgeons also demonstrated significantly higher intraobserver reliability as compared to the entire cohort (0.897 *vs.* 0.815; P<0.0001). Sub-analysis for agreement by SINS score demonstrated significantly higher intraobserver agreement among radiologists as compared to radiation oncologists (0.673 *vs.* 0.621; P=0.009), but worse interobserver reliability (0.452 *vs.* 0.509; P=0.01). As compared to the entire cohort, neither radiation oncologists (P=0.12), nor radiologists (P=0.30) demonstrated significant differences in interobserver reliability. Additionally, radiation oncologists did not show a significant difference from the overall cohort in terms of intraobserver reliability (P=0.35). Radiologists had significantly greater intraobserver reliability (0.673 *vs.* 0.636; P=0.006).

Discussion

The core features of an effective diagnostic test are consistency across observations, reliability across observers,

Table 4 Pooled estimates of intraobserver and interobserver reliability

Assessment	Studies	No. of Obs	Intraobserver reliability			Studies	No. of Obs	Interobserver reliability		
			Est	90% CI				Est	95% CI	
				LL	UL				LL	UL
All observers										
SINS score	(19-22)	162	0.815	0.661	0.969	(19-22)	162	0.673	0.227	1.12
SINS category	(9,10,19)	164	0.636	0.484	0.789	(9,10,19)	153	0.475	0.188	0.762
Location	(21,22)	49	0.882	0.756	1.01	(21-23)	66	0.831	0.643	1.02
Pain	(21,22)	49	0.836	0.782	0.889	(21-23)	66	0.694	0.340	1.05
Bone lesion quality	(21,22)	49	0.552	0.477	0.628	(21-23)	66	0.290	0.147	0.433
Alignment	(21,22)	49	0.612	0.560	0.664	(21-23)	66	0.467	0.360	0.579
Collapse	(21,22)	49	0.631	0.551	0.712	(21-23)	66	0.487	0.385	0.589
Posterior	(21,22)	49	0.610	0.514	0.707	(21-23)	66	0.453	0.289	0.617
Reliability of overall SINS score by specialty										
Spine surgeon	(21,22)	49	0.897	0.872	0.921	(21,22)	49	0.919	0.759	1.08
Orthopedic surgeon	N/A	N/A	N/A	N/A	N/A	(19,20)	20	0.675	0.210	1.14
Reliability of SINS category by specialty										
Radiation oncologists	(10,19)	55	0.621	0.412	0.830	(10,19)	55	0.509	0.291	0.727
Radiologist	(9,19)	60	0.673	0.543	0.803	(9,19)	60	0.453	0.213	0.692

Est, pooled estimate; Obs, observers; SINS, spinal instability neoplastic score.

and the ability to diagnosis the feature of interest. Additionally, to be clinically valuable, a diagnostic test must be able to alter or guide patient management. In this review we address the degree to which the Spinal Instability Neoplastic Score meets the first two criteria—consistency across observations (intraobserver reliability) and reliability across observers (interobserver reliability). This systematic review included 7 publications for both qualitative study and meta-analysis. Overall, we identified the SINS score to have substantial interobserver reliability and near perfect intraobserver reliability according to the method of Landis and Cook (11). Intraobserver reliability for SINS category was also found to be substantial and interobserver reliability, as with SINS score, was slightly poorer, being moderate overall. Sub-analysis of SINS categories demonstrated intraobserver agreement to be best for location and pain and poorest for bone quality. Lastly, we found that intraobserver and interobserver agreement were significantly higher for spine surgeons as compared to the general observer population. This is consistent with the notions that: (I) consistency across measures improves with increased familiarity with the system, and (II) interobserver reliability improves with increasingly similar training/background across observers.

Though not tested here, prior literature suggests that the SINS score also demonstrates the third feature of an effective diagnostic test—the ability to accurately measure the outcome of interest. Defined clinically, mechanical instability of the spine is progressive destruction of the spine ultimately resulting in breakdown (e.g., compression fracture). Factors contributing to mechanical instability include lesion size (24-26) and extent of osteolysis (27,28). Previous work in cadaveric models has demonstrated that bone mineral density—a non-pathologic analog of the extent of osteolysis—is directly correlated with Young's modulus and the ability to withstand compressive forces such as the axial loading experienced with ambulation (29-32). Similarly, in cadaveric models of lytic lesions, the majority of authors have documented an association between defect size and failure strength (33-35). To this end, Whyne *et al.* found that vertebral defect size was the strongest predictor of overall vertebral strength (35). Both features are incorporated into the SINS system as the bone quality and vertebral involvement/collapse domains. Prior clinical studies have suggested the posterior elements may act as a modifier of stability, as their destruction reduces the extent of vertebral involvement required to place the vertebra at risk of collapse (25).

Other domains of the SINS score are derived from manifestations of spinal instability, specifically pain, alignment, and vertebral body collapse. As the vertebral body becomes increasingly involved by the tumor, compressive strength decreases, ultimately resulting in anterior and middle column failure. This creates a focal kyphosis or deformity, which has been documented in patients with metastatic disease of the spine followed with serial imaging (36). Anterior and middle column collapse also shifts the moment arm of the superior spine segment anteriorly, placing greater strain on the pedicles, which are relatively fixed at the facet joints. In cases with concomitant metastatic pedicle involvement, focal kyphosis may then result in pedicle lysis and subluxation at the level of the involved segment.

In all cases of vertebral body collapse, there is a stretching of the periosteum, which is exacerbated by movement. This stretching leads to activation of $CGRP^+$ nociceptive afferents that innervate the periosteum, giving rise to the sensation of mechanical pain (37-41). These afferents, as well as similar afferents may also be activated by large or aggressive lesions without concomitant vertebral body collapse, generating the oncologic pain seen in many patients. As SINS incorporates radiographic markers previously correlated with decreased structural rigidity in addition to clinical and radiographic signs of structural compromise, it can be reasonably concluded that SINS score correlates with the clinical outcome of interest.

Considering this, as well as the robustness of testing results suggested in our meta-analysis, SINS appears to possess the three core features of an effective diagnostic test. Extant literature also suggests that it possesses the fourth feature of a valuable diagnostic test—the ability to guide management—although the evidence for this is less substantial. Versteeg *et al.* reported a multi-institutional series of 1,509 patients with spinal metastases treated with palliative surgery or radiation (42). Although it was not explicitly used to determine if a case was operative, SINS score was significantly higher in the surgery group. Additionally, they found that after publication of the SINS system mean SINS score dropped for patients in both treatment arms, consistent with the notion that SINS may enable oncology specialists to diagnose impending or gross mechanical instability earlier and therefore shorten the time to referral for intervention. Along similar lines, Hussain *et al.* recently published a prospective series demonstrating the ability of SINS to identify patients likely to experience improvement in pain and disability following surgical stabilization of metastatic spine lesions (18). They reported that even controlling for neurological status, SINS score positively correlated with preoperative pain and walking scores on both the Brief Pain Inventory (BPI) and the MD Anderson Symptom Inventory (MDASI). A significant correlation was also seen between SINS and postoperative pain (e.g., improvement of worst pain score on the BPI and overall pain score on the MDASI). Because of this, several authors have recommended incorporating SINS into clinical management algorithms for patients with spinal metastases, such as the NOMS framework of Laufer *et al.* (5) or the LMNOP framework of Ivanishvili and Fourney (43). Additionally, several professional societies, including the American Academy of Orthopedic Surgeons and the American College of Radiology, have recommended the use of SINS for the assessment of instability in the metastatic spine (44). However, no studies to date have reported how SINS precisely translates into clinical decision making in practice (i.e., what score should be used as the precise cutoff for the decision to stabilize?). Such a study would contribute substantially to the integration of SINS into clinical practice.

Conclusions

Here we report the results of a systematic review and meta-analysis of the extant medical literature describing the intraobserver and interobserver reliability of the SINS System for the diagnosis of mechanical instability in the metastatic spine. The results of our meta-analysis suggest the SINS score is highly reliable both within and across observers. Additionally, the degree of reliability seems to increase with increased clinical exposure to metastatic spine disease. Though it can be useful in guiding clinical management, additional data are needed to delineate a more precise cutoff in determining the need for stabilization in metastatic lesions of the spine.

Acknowledgments

None.

Footnote

Conflicts of Interest: ML Goodwin: Consultant for ROM3, Augmedics; DM Sciubba: Consultant for Orthofix, Globus, K2M, Medtronic, Stryker, Baxter. The other authors have no conflicts of interest to declare.

References

1. Budczies J, von Winterfeld M, Klauschen F, et al. The landscape of metastatic progression patterns across major human cancers. Oncotarget 2015;6:570-83.
2. Fornasier VL, Horne JG. Metastases to the vertebral column. Cancer 1975;36:590-4.
3. Klimo P Jr, Thompson CJ, Kestle JRW, et al. A meta-analysis of surgery versus conventional radiotherapy for the treatment of metastatic spinal epidural disease. Neuro Oncol 2005;7:64-76.
4. Barzilai O, Fisher CG, Bilsky MH. State of the Art Treatment of Spinal Metastatic Disease. Neurosurgery 2018;82:757-69.
5. Laufer I, Rubin DG, Lis E, et al. The NOMS framework: approach to the treatment of spinal metastatic tumors. Oncologist 2013;18:744-51.
6. Pope MH, Panjabi M. Biomechanical definitions of spinal instability. Spine 1985;10:255-6.
7. Fisher CG, DiPaola CP, Ryken TC, 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.
8. Joseph W. Rudmin. Calculating the Exact Pooled Variance James Madison University, 2010.
9. 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;203:869-74.
10. 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.
11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
12. Bilsky M, Fischer CG, Gokaslan ZL, et al. The Spinal Instability Neoplastic Score (SINS): An Analysis of Reliability and Validity from the Spine Oncology Study Group. Int J Radiat Oncol Biol Phys 2010;78:S263.
13. Hussain I, Barzilai O, Reiner A, et al. Patient-Reported Outcomes After Surgical Stabilization of Spinal Tumors: Symptom-Based Validation of the Spinal Instability Neoplastic Score (SINS) and Surgery. Spine J 2018;18:261-7.
14. Reynolds J, Fisher C, Boriani S, et al. Spinal Instability Neoplastic Score: An Analysis of Reliability and Validity in Radiologists. Eur Spine J 2013;22:S703.
15. Sahgal A, Schouten R, Versteeg A, et al. A Multi-institutional Study Evaluating the Reliability of the Spinal Instability Neoplastic Score (SINS) Among Radiation Oncologists for Spinal Metastases. Int J Rad Onc Biol Phys 2014;90:S350.
16. Spiess M, Hnenny L, Fourney D. Spinal instability neoplastic score (SINS) reliability analysis in spine residents and fellows in orthopedics and neurosurgery. Can J Neurol Sci 2014;40:S14.
17. Versteeg A, Oner C, Fisher C, et al. Validation of the Spinal Instability Neoplastic Score (SINS); A Retrospective Analysis. Eur Spine J 2014;23:S480.
18. Hussain I, Barzilai O, Reiner A, et al. Patient-Reported Outcomes After Surgical Stabilization of Spinal Tumors: Symptom-Based Validation of the Spinal Instability Neoplastic Score (SINS) and Surgery. Spine J 2018;18:261-7.
19. Arana E, Kovacs FM, Royuela A, et al. Spine Instability Neoplastic Score: agreement across different medical and surgical specialties. Spine J 2016;16:591-9.
20. Campos M, Urrutia J, Zamora T, et al. The Spine Instability Neoplastic Score: an independent reliability and reproducibility analysis. Spine J 2014;14:1466-9.
21. 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;29:3072-7.
22. Fox S, Spiess M, Hnenny L, et al. Spinal Instability Neoplastic Score (SINS): Reliability Among Spine Fellows and Resident Physicians in Orthopedic Surgery and Neurosurgery. Global Spine J 2017;7:744-8.
23. Teixeira WG, Coutinho PR, Marchese LD, et al. Interobserver agreement for the spine instability neoplastic score varies according to the experience of the evaluator.

Clinics (Sao Paulo) 2013;68:213-8.
24. Sahli F, Cuellar J, Pérez A, et al. Structural parameters determining the strength of the porcine vertebral body affected by tumours. Comput Methods Biomech Biomed Engin 2015;18:890-9.
25. Taneichi H, Kaneda K, Takeda N, et al. Risk factors and probability of vertebral body collapse in metastases of the thoracic and lumbar spine. Spine (Phila Pa 1976) 1997;22:239-45.
26. Weber MH, Burch S, Buckley J, et al. Instability and impending instability of the thoracolumbar spine in patients with spinal metastases: a systematic review. Int J Oncol 2011;38:5.
27. Alkalay R, Adamson R, Miropolsky A, et al. Female Human Spines with Simulated Osteolytic Defects: CT-based Structural Analysis of Vertebral Body Strength. Radiology 2018;288:436-44.
28. Alkalay RN, Harrigan TP. Mechanical assessment of the effects of metastatic lytic defect on the structural response of human thoracolumbar spine. J Orthop Res 2016;34:1808-19.
29. Briggs AM, Perilli E, Codrington J, et al. Subregional DXA-derived vertebral bone mineral measures are stronger predictors of failure load in specimens with lower areal bone mineral density, compared to those with higher areal bone mineral density. Calcif Tissue Int 2014;95:97-107.
30. Perilli E, Briggs AM, Kantor S, et al. Failure strength of human vertebrae: prediction using bone mineral density measured by DXA and bone volume by micro-CT. Bone 2012;50:1416-25.
31. Hipp JA, Rosenberg AE, Hayes WC. Mechanical properties of trabecular bone within and adjacent to osseous metastases. J Bone Miner Res 1992;7:1165-71.
32. Lee CH, Landham PR, Eastell R, et al. Development and validation of a subject-specific finite element model of the functional spinal unit to predict vertebral strength. Proc Inst Mech Eng H 2017;231:821-30.
33. Ebihara H, Ito M, Abumi K, et al. A Biomechanical Analysis of Metastatic Vertebral Collapse of the Thoracic Spine. Spine (Phila Pa 1976) 2004;29:994-9.
34. Dimar JR, Voor MJ, Zhang YM, et al. A human cadaver model for determination of pathologic fracture threshold resulting from tumorous destruction of the vertebral body. Spine (Phila Pa 1976) 1998;23:1209-14.
35. Whyne CM, Hu SS, Lotz JC. Burst fracture in the metastatically involved spine: development, validation, and parametric analysis of a three-dimensional poroelastic finite-element model. Spine (Phila Pa 1976) 2003;28:652-60.
36. Asdourian PL, Mardjetko S, Rauschning W, et al. An evaluation of spinal deformity in metastatic breast cancer. J Spinal Disord 1990;3:119-34.
37. Hiasa M, Okui T, Allette YM, et al. Bone Pain Induced by Multiple Myeloma Is Reduced by Targeting V-ATPase and ASIC3. Cancer Res 2017;77:1283-95.
38. Mach DB, Rogers SD, Sabino MC, et al. Origins of skeletal pain: sensory and sympathetic innervation of the mouse femur. Neuroscience 2002;113:155-66.
39. Jimenez-Andrade JM, Mantyh WG, Bloom AP, et al. A phenotypically restricted set of primary afferent nerve fibers innervate the bone versus skin: therapeutic opportunity for treating skeletal pain. Bone 2010;46:306-13.
40. Wakabayashi H, Wakisaka S, Hiraga T, et al. Decreased sensory nerve excitation and bone pain associated with mouse Lewis lung cancer in TRPV1-deficient mice. J Bone Miner Metab 2018;36:274-85.
41. Yoneda T, Hiasa M, Nagata Y, et al. Acidic microenvironment and bone pain in cancer-colonized bone. Bonekey Rep 2015;4:690.
42. Versteeg AL, van der Velden JM, Verkooijen HM, et al. The Effect of Introducing the Spinal Instability Neoplastic Score in Routine Clinical Practice for Patients With Spinal Metastases. Oncologist 2016;21:95-101.
43. Ivanishvili Z, Fourney DR. Incorporating the Spine Instability Neoplastic Score into a Treatment Strategy for Spinal Metastasis: LMNOP. Global Spine J 2014;4:129-36.
44. Versteeg AL, Verlaan J, Sahgal A, et al. The Spinal Instability Neoplastic Score: Impact on Oncologic Decision-Making. Spine (Phila Pa 1976) 2016;41 Suppl 20:S231-7.

Cite this article as: Pennington Z, Ahmed AK, Cottrill E, Westbroek EM, Goodwin ML, Sciubba DM. Intra- and interobserver reliability of the Spinal Instability Neoplastic Score system for instability in spine metastases: a systematic review and meta-analysis. Ann Transl Med 2019;7(10):218. doi: 10.21037/atm.2019.01.65