



Inter-rater reliability between musculoskeletal radiologists and orthopedic surgeons on computed tomography imaging features of spinal metastases

L. Khan MD, G. Mitera MBA,* L. Probyn MD,* M. Ford MD,* M. Christakis MD,* J. Finkelstein MD,* A. Donovan MD,* L. Zhang PhD,* L. Zeng,* J. Rubenstein MD,* A. Yee MD,* L. Holden MRTT,* and E. Chow MBBS**

ABSTRACT

Introduction

The primary objective of this pilot study was to examine the inter-rater reliability in scoring the computed tomography (CT) imaging features of spinal metastases in patients referred for radiotherapy (RT) for bone pain.

Methods

In a retrospective review, 3 musculoskeletal radiologists and 2 orthopedic spinal surgeons independently evaluated CT imaging features for 41 patients with spinal metastases treated with RT in an outpatient radiation clinic from January 2007 to October 2008. The evaluation used spinal assessment criteria that had been developed in-house, with reference to

- osseous and soft tissue tumour extent,
- presence of a pathologic fracture,
- severity of vertebral height loss, and
- presence of kyphosis.

The Cohen kappa coefficient between the two specialties was calculated.

Results

Mean patient age was 69.2 years (30 men, 11 women). The mean total daily oral morphine equivalent was 73.4 mg. Treatment dose-fractionation schedules included 8 Gy/1 ($n = 28$), 20 Gy/5 ($n = 12$), and 20 Gy/8 ($n = 1$). Areas of moderate agreement in identifying the CT imaging appearance of spinal metastasis included extent of vertebral body involvement ($\kappa = 0.48$) and soft-tissue component ($\kappa = 0.59$). Areas of fair agreement included extent of pedicle involvement ($\kappa = 0.28$), extent of lamina involvement ($\kappa = 0.35$), and presence of pathologic fracture ($\kappa = 0.20$). Areas of poor agreement

included nerve-root compression ($\kappa = 0.14$) and vertebral body height loss ($\kappa = 0.19$).

Conclusions

The range of agreement between musculoskeletal radiologists and orthopedic surgeons for most spinal assessment criteria is moderate to poor. A consensus for managing challenging vertebral injuries secondary to spinal metastases needs to be established so as to best triage patients to the most appropriate therapeutic modality.

KEY WORDS

Computed tomography, metastases, inter-rater reliability, spine

1. INTRODUCTION

On autopsy, 70%–80% of cancer patients show bone metastases¹. The most common site of bony metastasis is the spine², and associated back pain can result from tumour and mechanical causes. Tumour-related pain may be caused by inflammatory mediators, tumour stretching the periosteum of the vertebral body, and nerve-root compression³; mechanical pain results from structural abnormalities of the spine, such as a pathologic compression fracture. The most effective sequence of interventions—that is, the use of radiotherapy (RT) or minimally invasive surgical procedures—depends on reliable evaluation of metastatic spine involvement and its features^{4–6}.

Computed tomography (CT) is the imaging modality most commonly used for evaluating the bony spine. However, despite continuing advancement in CT image resolution and better visibility of tumours, variability in interpretation of the target volume remains a major source of error^{7,8}. Bowden *et al.* found that application of a delineation protocol improved accuracy in identifying target volume. Their improved protocol includes guidelines concerning

level and window settings and tumour identification by a diagnostic radiologist⁹.

The decision to proceed to surgery is often made from an evaluation of CT imaging by the orthopedic spinal surgeon (OSS) involved. In 1989, Mirels proposed an innovative image-based scoring system for impending pathologic fractures of long bones¹⁰; that system has been highly utilized since its conception. A similar prognostic tool, developed within a multi-disciplinary setting, for spinal compromise secondary to metastatic disease is not currently available.

The purpose of the present pilot study was to assess reliability in the scoring of CT imaging features between musculoskeletal radiologists (MSKS) and OSSs.

2. METHODS

We retrospectively reviewed 41 patients with spinal metastases who were receiving RT in an outpatient palliative clinic at a tertiary care hospital from January 2007 to October 2008. Given the retrospective nature of this study, the assessment was performed using the available CT imaging from routine RT simulation (3-mm slices). Features of the CT images were independently evaluated by 3 MSKS and 2 OSSs, using in-house spinal assessment criteria that included features that both expert groups thought important to capture (Appendix A). The spinal assessment criteria included

- radiated site, including cervical, thoracic, lumbar, and sacral spine;
- extent of tumour involvement;
- type of lesion (that is, sclerotic, lytic, mixed);
- presence of pathologic fracture;
- height loss;
- column involvement;
- soft-tissue component;
- nerve-root compression; and
- kyphosis.

The MSK scoring was considered the clinical standard.

2.1 Statistical Analysis

Descriptive statistics are expressed as means and standard deviations for quantitative variables and as proportions for qualitative variables. The percentage agreement between MSKS and OSSs was calculated for each spinal metastasis assessment criterion. The weighted Cohen kappa coefficient (κ) was also calculated after adjusting for weighting information to test for the percentage agreement between the two specialties at the 95% confidence interval. The weight for calculation was considered using a binary variable of cancer seen (1 = Yes, 0 = No), because some cancers were seen only by the MSK group. Primary cancer type was recorded based on the pathology report and captured in the demographic data. Some weighted kappa values were not calculated because

of the lower number of cells in the cross table. A kappa value of 1 implies perfect agreement; values less than 1 imply less perfect agreement. These were the agreement categories used in the study¹¹:

- <0.20: Poor agreement
- 0.20–0.40: Fair agreement
- 0.41–0.60: Moderate agreement
- 0.61–0.80: Good agreement
- 0.81–1.00: Very good agreement

All statistical analyses were conducted using the Statistical Analysis Software application (version 9.2: SAS Institute, Cary, NC, USA) for Windows (Microsoft, Redmond, WA, USA).

3. RESULTS

3.1 Baseline Patient Demographics

The mean age of the 41 study patients (11 women, 31 men) was 69.2 ± 12.3 years. Their mean total daily oral morphine equivalent was 73.4 mg. The primary cancer sites were mainly prostate ($n = 21$, 51%), breast ($n = 9$, 22%), and lung ($n = 7$, 17%). The most frequently treated site was the lumbar spine ($n = 20$, 49%), followed by the thoracic ($n = 13$, 34%), the cervical ($n = 6$, 15%), and the sacral ($n = 1$, 2%). Treatment dose–fractionation schedules included 8 Gy/1 ($n = 28$, 68%), 20 Gy/5 ($n = 12$, 29%), and 20 Gy/8 ($n = 1$, 2%). Table 1 lists complete demographic details for the patients.

3.2 Assessment Criteria

Moderate agreement between the specialties was observed for

- extent of soft-tissue component (87.2%, $\kappa = 0.59$),
- extent of vertebral body involvement (56.8%, $\kappa = 0.48$),
- pedicle involvement (71.8%, $\kappa = 0.42$),
- laminar involvement (74.4%, $\kappa = 0.43$),
- right posterior column (71.8%, $\kappa = 0.44$), and
- left posterior column (71.8%, $\kappa = 0.41$).

Fair agreement was observed for

- extent of laminar involvement (85.7%, $\kappa = 0.35$),
- type of pathologic fracture [burst, wedge, other (73%, $\kappa = 0.20$)],
- right anterior column (69.2%, $\kappa = 0.30$),
- right middle column (74.4%, $\kappa = 0.24$), and
- extent pedicle involvement (64.3%, $\kappa = 0.285$).

Agreement was poor with regard to quantifying

- extent of vertebral body involved in the fracture (33.3%),

TABLE I Patient demographics

Variable	Value
Patients (<i>n</i>)	41
Age (years)	
Mean	69.2±12.3
Median (range)	70 (46–87)
Sex (women/men)	11/30
Primary cancer sites [<i>n</i> (%)]	
Prostate	21 (51.2)
Breast	9 (22)
Lung	7 (17.1)
Bladder	2 (4.9)
Colon	1 (2.4)
Unknown	1 (2.4)
Dose/fractions [<i>n</i> (%)]	
800 cGy/1	28 (68.3)
2000 cGy/5	12 (29.3)
2000 cGy/8	1 (2.4)
Radiation vertebral site [<i>n</i> (%)]	
Lumbar	20 (48.8)
Thoracic	14 (34.1)
Cervical	6 (14.7)
Sacral	1 (2.4)

- height loss (75.8%, $\kappa = 0.19$),
- nerve-root component (76.9%, $\kappa = 0.14$), and
- left anterior column (66.7%, $\kappa = 0.15$).

Agreement was 94.7% for kyphosis, 97.5% for vertebral body involvement (yes/no), and 84.6% for the type of lesion (lytic, sclerotic, or mixed). For these last three criteria, the Cohen kappa could not be calculated because the sample size was almost negligible. Tables II and III respectively set out the percentage agreement between the MSKS and OSSS and the Cohen kappa coefficient for inter-rater agreement on the spinal metastases assessment criteria. Overall, high-percentage agreement was observed in most areas, and the inter-rater agreement between the two specialties ranged from moderate to poor (Figure 1).

4. DISCUSSION

Our study investigated the inter-rater reliability between OSS and MSK specialists in the assessment of disruption in the bony architecture secondary to spinal metastases. Simulation CT scans with 3-mm slices were used in the assessments. The following findings are salient:

- Moderate agreement in differentiating pedicle or lamina involvement

TABLE II Percentage agreement between musculoskeletal radiologists and orthopedic spinal surgeons

Assessment criterion	Agreement (%)
Extent of tumour involvement	
Vertebral body	97.5
Vertebral involvement	56.8
Posterior elements	
Pedicle	71.8
Pedicle involvement	64.3
Pedicle reason	87.5
Laminar	74.4
Laminar involvement	85.7
Laminar reason	100.0
Type of lesion	84.6
Pathologic fracture	73.0
Burst	100.0
Wedge	33.3
Other	66.7
Invading >50% of vertebral body	33.3
Superior endplate	33.3
Inferior endplate	33.3
Height loss	75.8
Column involvement	
Left anterior (A1)	66.7
Right anterior (A2)	69.2
Left middle (M1)	74.4
Right middle (M2)	69.2
Left posterior (P1)	71.8
Right posterior (P2)	71.8
Soft-tissue component	87.2
Nerve-root component	76.9
Kyphosis	94.7

- Poor agreement for the degree of pedicle or lamina involvement
- Fair agreement regarding fracture type
- Poor agreement regarding vertebral height loss
- Poor agreement for indentifying nerve-root compression

The poor agreement for lamina, pedicle, and anterior or posterior column involvement is most likely a result of the known difficulty in quantifying metastatic bone disease. To effectively quantify metastatic tumour involvement in the spine, accurate segmentation of the vertebra is required. Manual segmentation can be accurate, but involves extensive and time-consuming user interaction¹². Hardisty *et al.* proposed an algorithm that allows for semi-automated quantification of bone involvement by tumour; however, their method is still time-consuming, useful only in the hands of an experienced user, and not widely available¹³.

INTER-RATER RELIABILITY STUDY FOR SPINAL METASTASES

TABLE III Cohen kappa coefficient for inter-rater agreement between musculoskeletal radiologists and orthopedic spinal surgeons on spinal metastases assessment criteria

Assessment criterion	Inter-rater agreement		
	Agreement	Cohen kappa coefficient	95% Confidence limits
Extent of tumour involvement			
Vertebral body	—	—	—
Vertebral % involvement	Moderate	0.4777	(0.2814, 0.6740)
Posterior elements			
Pedicle	Moderate	0.4226	(0.1354, 0.7098)
Pedicle >50% or ≤50% involvement	Fair	0.2853	(-0.0086, 0.5793)
Pedicle reason	Poor	-0.0667	(-0.1589, 0.0255)
Laminar	Moderate	0.4298	(0.1390, 0.7207)
Laminar >50% or ≤50% involvement	Fair	0.3568	(0.0507, 0.6630)
Laminar reason	—	—	—
Type of lesion			
Pathologic fracture	Fair	0.2060	(-0.1482, 0.5602)
Height loss	Poor	0.1902	(-0.1985, 0.5789)
Column involvement			
Left anterior (A1)	Poor	0.1536	(-0.1758, 0.4830)
Right anterior (A2)	Fair	0.3097	(-0.0010, 0.6204)
Left middle (M1)	Fair	0.2412	(-0.0812, 0.5637)
Right middle (M2)	Poor	0.1959	(-0.1361, 0.5278)
Left posterior (P1)	Moderate	0.4147	(0.1278, 0.7017)
Right posterior (P2)	Moderate	0.4407	(0.1786, 0.7027)
Soft-tissue component	Moderate	0.5979	(0.2967, 0.8992)
Nerve-root component	Poor	0.1418	(-0.1106, 0.3942)
Kyphosis	—	—	—

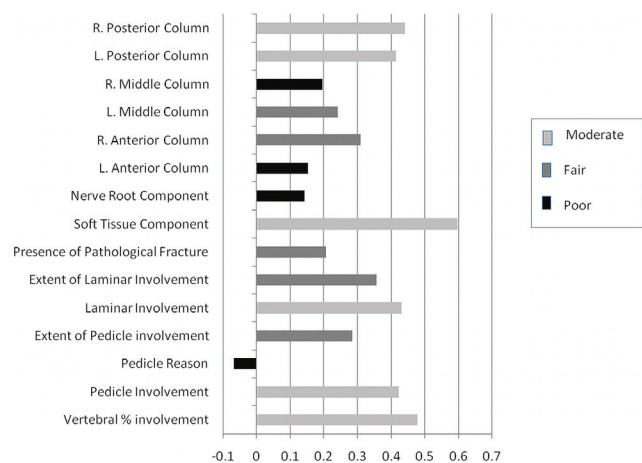


FIGURE 1 Cohen kappa coefficient for inter-rater agreement between musculoskeletal radiologists and orthopedic spinal surgeons on spinal metastases assessment criteria. R = right; L = left.

Differentiating between wedge and burst in a pathologic fracture was also poor, but such differentiation is somewhat of a gray area. The definition of burst is based on retropulsion of the posterior cortex.

With a wedge fracture, the appearance of the posterior cortex can be similar if the tumour extends posterior.

Agreement was also poor with regard to height loss. That lack of agreement raises concerns, because patients with vertebral collapse have been shown to benefit from surgical intervention^{4,6}. Walraevens *et al.*¹⁴ showed that, for precise identification of height loss, objective scoring systems are required. These scoring systems often depend on experience and the discipline involved.

Another area of poor agreement was the identification of nerve-root compression. Compared with CT, magnetic resonance imaging is well known to be a more sensitive modality for evaluating the spinal canal and nerve roots, and that difference could perhaps explain the level of discordance in scores based on CT imaging^{15,16}.

The identified areas of poor agreement are central to clinical decision-making, and thus highlight the need for objective measures to quantify disease, to validate clinical outcome, to contribute to the efficiency of clinical trials, and to raise the degree of certainty for clinicians attempting to correlate interval change with true change in the

clinical status of the patient¹⁷. The manner in which different clinicians “see” the tumour burden in a vertebra forms their perception of the clinical issues of stability and prognosis. If the tumour is variably measured, then clinical judgments will be expected to similarly vary. The present study may serve as early investigational step in determining the need for a prognostic tool to evaluate metastases to the spine. Our findings might be important in formulating consensus-based protocols for a multidisciplinary approach to managing challenging vertebral injuries secondary to spinal metastases.

Our study is limited by the sample size and the diagnostic quality of the imaging. Specifically, we note that the kappa values were influenced by the sample size rather than by variance in the interpretation between the two specialities¹¹. Validation of this study in a larger cohort of patients undergoing either surgery or radiation therapy should be considered for the future. Lack of agreement in the scoring of CT imaging features could prove to be a critical factor in therapeutic decision-making. A standardized method of characterizing spinal metastases with explicit guidelines would be helpful in triaging patients to the most appropriate treatments.

5. ACKNOWLEDGMENTS

We thank the Michael and Karyn Goldstein Cancer Research Fund, Sarah Campos, Shaelyn Culleton, and Stacy Yuen for assistance.

6. CONFLICT OF INTEREST DISCLOSURES

The authors have no conflicts of interest related to this manuscript to declare.

7. REFERENCES

1. Tubiana-Hulin M. Incidence, prevalence and distribution of bone metastases. *Bone* 1991;12(suppl 1):S9–10.
2. Bilsky MH, Lis E, Raizer J, Lee H, Boland P. The diagnosis and treatment of metastatic spinal tumour. *Oncologist* 1999;4:459–69.
3. Gokaslan ZL, York JE, Walsh GL, *et al.* Transthoracic vertebrectomy for metastatic spinal tumors. *J Neurosurg* 1998;89:599–609.
4. Fournay DR, Schomer DF, Nader R, *et al.* Percutaneous vertebroplasty and kyphoplasty for painful vertebral body fractures in cancer patients. *J Neurosurg* 2003;98(suppl 1):21–30.
5. Weill A, Chiras J, Simon JM, Rose M, Sola-Martinez T, Enkaoua E. Spinal metastases: indications for and results of percutaneous injection of acrylic surgical cement. *Radiology* 1996;199:241–7.
6. Chow E, Holden L, Danjoux C, *et al.* Successful salvage using percutaneous vertebroplasty in cancer patients with painful spinal metastases or osteoporotic compression fractures. *Radiother Oncol* 2004;70:265–7.
7. Weiss E, Hess CF. The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy. *Strahlenther Onkol* 2003;179:21–30.
8. Roy AE, Wells P. Volume definition in radiotherapy planning for lung cancer: how the radiologist can help. *Cancer Imaging* 2006;7:116–23.
9. Bowden P, Fisher R, Mac Manus M, *et al.* Measurement of lung tumor volumes using three-dimensional computer planning software. *Int J Radiat Oncol Biol Phys* 2002;53:566–73.
10. Mirels H. Metastatic disease in long bones. A proposed scoring system for diagnosing impending pathologic fractures. *Clin Orthop Relat Res* 1989;:256–64.
11. Altman DG. *Practical Statistics for Medical Research*. London, UK: Chapman and Hall; 1991.
12. Whyne C, Hardisty M, Wu F, *et al.* Quantitative characterization of metastatic disease in the spine. Part II. Histogram-based analyses. *Med Phys* 2007;34:3279–85.
13. Hardisty M, Gordon L, Agarwal P, Skrinkas T, Whyne C. Quantitative characterization of metastatic disease in the spine. Part I. Semiautomated segmentation using atlas-based deformable registration and the level set method. *Med Phys* 2007;34:3127–34.
14. Walraevens J, Liu B, Meersschaert J, *et al.* Qualitative and quantitative assessment of degeneration of cervical intervertebral discs and facet joints. *Eur Spine J* 2009;18:358–69.
15. Kardamakis D, Vassiliou V, Chow E, eds. *Bone Metastases: A Translational and Clinical Approach*. Series: Cancer Metastasis—Biology and Treatment. Dordrecht, Netherlands: Springer; 2009.
16. Carrino JA, Lurie JD, Tosteson AN, *et al.* Lumbar spine: reliability of MR imaging findings. *Radiology* 2009;250:161–70.
17. Furlan JC, Fehlings MG, Massicotte EM, *et al.* A quantitative and reproducible method to assess cord compression and canal stenosis after cervical spine trauma: a study of interrater and intrarater reliability. *Spine* 2007;32:2083–91.

Correspondence to: Gunita Mitera, Department of Radiation Therapy, Odette Cancer Centre, Sunnybrook Health Sciences Centre, 2075 Bayview Avenue, Toronto Ontario M4N 3M5.

E-mail: Gunita.Mitera@sunnybrook.ca

* Bone Metastases Site Group, Odette Cancer Centre, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, ON.

**APPENDIX A
SPINAL METASTASES ASSESSMENT
CRITERIA**

1. Radiated site:

- Cervical Spine Level _____ (specify)
- Thoracic Spine Level _____ (specify)
- Lumbar Spine Level _____ (specify)

2. Extent of tumor involvement:

- Vertebral body only = _____% involvement

- Posterior elements = _____% involvement
- Pedicle
- Due to tumor
- Due to fracture
- Due to both

- Laminar = _____% involvement
- Due to tumor
- Due to fracture
- Due to both

3. Type of lesion:

- Sclerotic
- Lytic
- Mixed

4. Pathological Fracture:

- Yes If yes: Burst
- No Wedge
- Other

- Invading > 50% of vertebral body

- Superior Endplate
- Inferior Endplate

5. Height Loss

- Yes = _____% of vertebral body collapse
- No

6. Column Involvement*

- L. Anterior (A1) R. Anterior (A2)
- L. Middle (M1) R. Middle (M2)
- L. Posterior (P1) R. Posterior (P2)

*Please list from most affected zone to least:

7. Soft Tissue Component

- Yes
- No

8. Nerve Root Compression

- Yes
- No

9. Kyphosis

- Yes = _____° (absolute degrees)
- No