Radiotherapy and Oncology 115 (2015) 135–140

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Metastatic spine disease

# Agreement in the assessment of metastatic spine disease using scoring systems

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

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

ARTICLE INFO

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

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

ABSTRACT

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


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

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

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

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

http://dx.doi.org/10.1016/j.radonc.2015.03.016
0167-8140/

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

## Methods

### *Study design and participants*

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

### *Selection of hospital departments and clinicians*

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

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

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

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

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

### *Patients and images selection*

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

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

### *Procedure*

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

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

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

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

### *Statistical analysis*

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

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

**Table 1**
Sample characteristics.

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

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

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

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

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

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

## Results

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## Discussion

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

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

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

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

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

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

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

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

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

## Conflicts of interest

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

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

## Funding

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

## Acknowledgements

Thank you to Prof David Moratal for his collaboration.

## Appendix A. Supplementary data

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

## References

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

assessment of instability secondary to spinal metastases. AJR Am J Roentgenol 2014;203:869–74. http://dx.doi.org/10.2214/AJR.13.12269.

[23] Bosch X, Moreno P, Ríos M, Jordán A, López-Soto A. Comparison of quick diagnosis units and conventional hospitalization for the diagnosis of cancer in Spain: a descriptive cohort study. Oncology 2012;83:283–91. http://dx.doi.org/10.1159/000341658.

[24] Fleissig A, Jenkins V, Catt S, Fallowfield L. Multidisciplinary teams in cancer care: are they effective in the UK? Lancet Oncol 2006;7:935–43. http://dx.doi.org/10.1016/S1470-2045(06)70940-8.

[25] Keating NL, Landrum MB, Lamont EB, Bozeman SR, Shulman LN, McNeil BJ. Tumor boards and the quality of cancer care. J Natl Cancer Inst 2013;105:113–21. http://dx.doi.org/10.1093/jnci/djs502.

[26] Dea N, Fisher CG. Evidence-based medicine in metastatic spine disease. Neurol Res 2014;36:524–9. http://dx.doi.org/10.1179/1743132814Y.0000000365.

[27] Bollen L, van der Linden YM, Pondaag W, Fiocco M, Pattynama BPM, Marijnen C, et al. Prognostic factors associated with survival in patients with symptomatic spinal bone metastases: a retrospective cohort study of 1,043 patients. Neuro Oncol 2014;16:991–8. http://dx.doi.org/10.1093/neuonc/not318.

[28] Krishnan MS, Epstein-Peterson Z, Chen Y-H, Tseng YD, Wright AA, Temel JS, et al. Predicting life expectancy in patients with metastatic cancer receiving palliative radiotherapy: the TEACHH model. Cancer 2014;120:134–41. http://dx.doi.org/10.1002/cncr.28408.

[29] Lim C, Cheung MC, Franco B, Dharmakulaseelan L, Chong E, Iyngarathasan A, et al. Quality improvement: an assessment of participation and attitudes of medical oncologists. J Oncol Pract 2014;10:e408–14. http://dx.doi.org/10.1200/JOP.2014.001515.