Mirels' Rating for Humerus Lesions is Both Reproducible and Valid

Andrew R. Evans MD, John Bottros MD,
William Grant EdD, Benjamin Y. Chen BS,
Timothy A. Damron MD

Abstract Mirels' rating system is commonly used to predict risk of fracture in patients with metastatic bone lesions to long bones, but it has not been independently validated for use in humeral bone lesions. We asked whether this system was a valid and reproducible instrument for predicting impending pathologic fractures in the humerus. We presented 17 case histories and plain radiographs of 16 patients with humeral metastases through a web-based survey to 39 physicians with varying training and experience. Participants scored each case using Mirels' criteria and provided a fracture prediction, which was compared with actual outcome in the subset of 12 patients with three fractures not treated prophylactically. Using Mirels' definition of impending pathologic fracture (nine points or greater), the sensitivity and specificity for determining the likelihood of pathologic humeral fracture were 14.5% and 82.9%, respectively. When we used seven or more points as the definition of impending pathologic humeral fracture, sensitivity improved to 81% but specificity was reduced to 32%. Kappa analysis suggested moderate reproducibility across groups for prediction of pathologic fracture. The Mirels rating system for humeral lesions is reproducible and valid, but low specificity at acceptable sensitivity levels as reported remains a problem as for femoral lesions.

Level of Evidence: Level III, diagnostic study. See the Guidelines for Authors for a complete description of levels of evidence.

A. R. Evans, W. Grant, B. Y. Chen, T. A. Damron
SUNY Upstate Medical University, Syracuse, NY, USA

J. Bottros
Cleveland Clinic Foundation, Cleveland, OH, USA

T. A. Damron (✉)
Institute for Human Performance, Musculoskeletal Science Research Center, 505 Irving Avenue, Room 3117, Syracuse, NY 13210, USA
e-mail: damront@upstate.edu

Introduction

Multiple instruments and criteria for predicting the risk of pathologic fracture in skeletal lesions have been published in recent decades, but their validity and applicability to clinical situations remains largely untested [1, 3, 6, 9–12]. In 1989, Mirels [9] published a composite weighted scoring system for predicting the likelihood of pathologic fracture based on four variables believed to contribute to pathologic fracture risk: location, pain level, radiographic appearance, and size. When considered independently, these risk factors perform poorly when predicting pathologic fracture risk [7]. However, when used in aggregate, they reportedly identify impending pathologic fractures with a sensitivity of 91% and a specificity of 33% [7]. One subsequent study [2] suggested the Mirels scoring system is reproducible and valid when used by musculoskeletal oncologists, orthopaedic surgeons, residents, radiologists, and medical and radiation oncologists to predict risk of fracture in pathologic lesions of the femur.

There are considerable differences in load-bearing requirements between the upper and lower extremities, thereby potentially conferring a different fracture susceptibility profile to the humerus as compared with the femur. Given these differences, we suspected the responsiveness

and reliability of the Mirels scoring system, particularly as it defines impending pathologic fracture at nine or greater in the femur, may be altered in the humerus.

We therefore asked whether the Mirels scoring system remains a reproducible and valid instrument for identifying impending pathologic fractures of the humerus with greater accuracy than clinical judgment alone.

Materials and Methods

We invited 72 physicians to complete an online survey consisting of 17 anonymous case histories describing 16 patients with metastatic lesions of the humerus at initial presentation. Thirty-nine of the 72 physicians (43%) responded to the online survey with 31 physicians completing the entire survey. Physician respondents who completed the survey were grouped into one of five cohorts as defined by specialty and/or level of training: (1) orthopaedic oncologists (four); (2) orthopaedic attending faculty (six); (3) orthopaedic residents (15); (4) musculoskeletal radiologists (three); and (5) medical/radiation oncologists (three). The orthopaedic oncologist cohort consisted of fellowship-trained orthopaedic oncologists from across the United States, all of whom had extensive experience in treating patients with metastatic disease to long bones. Orthopaedic faculty included attendings at one of two institutions (one academic and one private) who treat predominantly adult patients and have some experience in treating patients with metastatic disease of long bones. The orthopaedic residents were in their postgraduate years 3, 4, or 5 of a 5-year residency program. The musculoskeletal radiology cohort consisted of radiologists with specialized fellowship training in musculoskeletal radiology. The medical and radiation oncologists were faculty at academic institutions with extensive experience in treating patients with metastatic cancer.

Selection criteria for the physicians invited to participate differed according to cohort. For the orthopaedic oncologists, we chose to initially invite those who participated in the original study that validated the use of Mirels' [9] rating system in the femur. However, this was eventually expanded to include others who had expressed verbal interest or who had been active participants in other academic aspects of orthopaedic oncology. For the other cohorts, the physicians invited were predominately at institutions affiliated with those of the first author (TAD). The attending orthopaedic surgeons invited were involved predominantly in adult reconstructive orthopaedics treating patients with metastatic disease. For orthopaedic residents, only those in the PGY-3 year or above were invited to participate. Invitations were extended to residents at two institutions. For musculoskeletal radiologists, only those

whose primary practice involved musculoskeletal radiology and who worked on a regular basis with an orthopaedic oncologist were invited. For medical and radiation oncologists, only those who dealt primarily with adults and whose practices included patients with metastatic disease and myeloma were invited. No attempt was made to select physicians according to a similar treatment philosophy.

None of the case series patients had pathologic fractures before presentation. One of the patients presented with bilateral humeral involvement, and therefore the case was divided such that survey recipients evaluated and scored each humerus independently. Sixteen of 17 cases were presented with two orthogonal plain radiographs of the involved humerus. The remaining case was presented with a single anteroposterior plain radiograph of the involved humerus. All radiographs were enlarged to actual size and resolution by clicking directly on the images. Through the online questionnaire, physician respondents were also supplied with direct hyperlinks to the original Mirels manuscript as well as to a table summarizing the Mirels scoring scheme (Table 1) [9].

All patients included in this case series presented before fracture or treatment for their humeral lesion and without reference to any subsequent treatment or outcome. Each case history in the online survey presented a uniform set of information describing variables believed important in the assessment of pathologic lesions and analysis of the associated risk of pathologic fracture (Fig. 1) (Appendix 1). Cases were selected to provide a spectrum of clinical and radiographic scenarios with primary diagnoses including multiple myeloma (five), breast carcinoma (three), lung carcinoma (three), prostate carcinoma (one), bladder carcinoma (one), ovarian carcinoma (one), squamous cell carcinoma (one), malignant fibrous histiocytoma (one), and melanoma (one). There were seven women and nine men. The mean patient age was 64.1 years (range, 42–86 years). Lesion location within the humerus was distributed among metaphyseal (six), metadiaphyseal (six), and diaphyseal (five) locations. Treatment included prophylactic stabilization of five humeral lesions that were considered impending pathologic fractures. The remaining 12 humeral lesions served as the natural history subgroup and were treated nonoperatively. Of these latter 12 cases, three went

Table 1. Mirels' scoring system [9]

Variable	Score		
	1	2	3
Pain	Mild	Moderate	Functional
Location	Upper extremity	Lower extremity	Peritrochanteric
Size	Less than ⅓	⅓ to ⅔	Greater than ⅔
Nature	Blastic	Mixed	Lytic

Age at presentation for upper extremity disease
Gender
Race
Primary diagnosis
Time since initial diagnosis
Distribution of disease
Duration of upper extremity symptoms
Pain character
Pain medication
Response to pain medication
Surgical treatments
Chemotherapy and/or radiation treatment history
Pain response to treatment modalities
Medical comorbidities
Reason for presentation

Fig. 1 We included 15 factors in each case presentation although the details varied.

on to fracture and nine did not fracture or require prophylactic stabilization throughout the duration of their followup to the time of their death. For the purposes of evaluating reproducibility, all 17 humeral cases were included in the analysis, but for the purposes of analysis of validity and fracture prediction, only the 12 cases not treated nonoperatively initially were included.

We asked each respondent to review the original Mirels [9] manuscript before completing the survey. Completion of the survey required each respondent to score each individual Mirels component (pain, location, size, and nature) and calculate the total score for each case. In addition, we asked each respondent to make a determination as to whether each lesion represented an impending pathologic fracture independent of the Mirels score, make a recommendation regarding whether to perform prophylactic stabilization, and to provide a rationale for recommended treatment, particularly if their recommendation did not correspond to that suggested in the Mirels [9] manuscript. Responses were then compared with actual outcome: (1) nonoperative treatment with subsequent pathologic fracture; and (2) nonoperative treatment without subsequent fracture. We excluded incomplete responses from analysis.

One of us (WG) performed the analysis for reproducibility across raters using SAS-PC for Windows v.9.1.3™ (SAS Institute, Cary, NC). Specificity, sensitivity, and receiver operating curve (ROC) analyses were conducted using MedCalc software version 9.3.8.0 (MedCalc Software, Mariakerke, Belgium.)

Kappa values were used to evaluate levels of interobserver agreement and concordance when comparing fracture prediction and actual outcome for respondents within each cohort and for all respondents. We performed kappa analysis for multiple raters using examiner experience categorical classifications for all 17 cases [5]. Kappa values of 0.8 to 1.0 were considered to have an almost perfect level of agreement, values from 0.6 to 0.8 to have a substantial level of agreement, values from 0.4 to 0.6 to have moderate agreement, values from 0.2 to 0.4 to have fair agreement, and values less than 0.2 to have poor agreement [8]. Kappa analysis was performed on the pain, location, and size components of the Mirels rating system, but because of inadequate completion rates, the kappa analysis could not be completed for the radiographic appearance component (eg, lytic, mixed, blastic) (see Table 2).

Outcome prediction and actual outcome data were used to generate ROC curves for the entire group of respondents as applied to all 17 cases and to the natural history subgroup (all 12 patients managed nonoperatively with or without subsequent fracture). We then generated ROC curves for each respondent cohort individually; the area under the ROC curve serves as an estimate of prediction accuracy. For each curve we calculated sensitivity and specificity profiles with corresponding p values, 95% confidence intervals, and likelihood ratios.

Results

We found the Mirels [9] rating system reproducible. The respondents' total Mirels' rating system scores ranged from 6 to 10 (standard deviation, 1.29). Kappa analysis across all respondent cohorts revealed fair interobserver agreement for pain and size, slight agreement for location, and

Table 2. Reproducibility analysis for individual Mirels score components

Category	Standard error	Kappa value	Overall kappa p value	Kendall's coefficient of concordance p value
Pain	0.009	0.305	< 0.0001	< 0.0001
Location	0.013	0.018	0.0823	0.02
Size	0.01	0.363	< 0.0001	< 0.0001
Nature	Incomplete	Incomplete	Incomplete	Incomplete
Impending fracture prediction	0.013	0.443	< 0.0001	< 0.0001

Table 3. Comparison of sensitivity and specificity for total Mirels scores

Mirels score	Sensitivity	Specificity	Positive likelihood ratio
6	94.5	12.1	1.08
7	81.4	32.1	1.20
8	44.1	61.4	1.14
9	14.5	82.9	0.85

Table 4. Comparison of cohorts' relative ability to predict fracture

Group	Odds ratio	Sensitivity	Specificity	aROC
Musculoskeletal radiologists	2.48	60.0	78.3	0.69
Attending orthopaedists	2.41	61.5	81.5	0.72
Orthopaedic residents	1.80	68.3	63.7	0.66
Musculoskeletal oncologists	0.95	66.7	57.9	0.62
Medical/radiation oncologists	0.66	55.6	48.8	0.52

aROC = adjusted receiver operating characteristic.

moderate agreement for overall prediction of impending pathologic fracture. We found interobserver agreement and concordance (F = 25.6; p < 0.0001) for pain, size, and overall prediction of pathologic fracture.

Validity was also substantiated. The area under the ROC curve for all cohorts combined was 0.55 (95% confidence interval [CI], 0.51–0.59; p = 0.057), demonstrating the Mirels [9] rating system identified impending fractures better than chance alone. Evaluation of the sensitivity and specificity profile indicated the best accuracy at a total Mirels score of seven with an associated sensitivity and specificity of 81% and 32%, respectively (Table 3). The sensitivity and specificity of using 9 or greater as the definition for impending pathologic fracture (as suggested in the original Mirels [9] manuscript) in the humerus were 14.5% and 82.9%, respectively. Use of the Mirels rating system for three of the groups of evaluators improved their ability to predict fracture when compared to clinical judgment alone (odds ratio 3.05, 95% CI 1.99–4.68 versus 1.58, CI 1.03–2.42). In the evaluation of each cohort independently, musculoskeletal radiologists, attending orthopaedists, and orthopaedic residents were each as a group able to predict fracture using the Mirels rating system better than by using their clinical judgment alone. Musculoskeletal oncologists were able to predict fracture almost equivalently using the Mirels rating system versus clinical judgment, and medical/radiation oncologists were unable to enhance fracture prediction using the Mirels rating system (Table 4) [9].

Discussion

The Mirels rating system was originally described in a group of patients who predominantly had femoral lesions from breast cancer [9]. More recently, independent validation of the use of Mirels' rating system for predicting fracture in lesions isolated to the femur from a variety of metastatic primaries and myeloma was evaluated [2]. However, to date, there had been no previous independent validation of this rating system for predicting fracture in lesions isolated to the femur. We therefore asked three questions regarding the use of the Mirels scoring system as an instrument for identifying impending pathologic fractures of the humerus. (1) Is it reproducible? (2) Is it valid? (3) Does it predict humeral fractures with greater accuracy than clinical judgment alone and for providing recommendations for treatment that minimize morbidity?

The primary limitation of this study was the low response rate from many cohorts, particularly orthopaedic oncologists, musculoskeletal radiologists, and medical/radiation oncologists. We could not include the medical/radiation oncologist cohort in the kappa analysis as a result of an insufficient number of responses to the survey. The medical/radiation oncologist cohort, in addition to having the lowest level of participation of any cohort, also demonstrated the greatest variability in responses, the least accuracy in predicting pathologic fractures, and the lowest sensitivity and specificity using the Mirels rating system [9]. Although we cannot be absolutely certain, it is possible an increase in participation by these cohorts may have altered the odds ratios, sensitivities, and specificities used to evaluate their relative ability to predict fracture and ability to determine the necessity of prophylactic stabilization. However, this trend suggests further education of medical and radiation oncologists in the evaluation of metastatic long bone lesions, and particularly in the use of instruments such as the Mirels rating system [9], is necessary because they are commonly involved in the initial discovery of pathologic bone lesions and play a key role in deciding when to refer patients to orthopaedic oncologists for treatment.

In the published literature on Mirels' rating system, the original definition of determination of impending pathologic fracture utilized as a 'cut-off' a cumulative score of nine points or greater for all long bone lesions, and this was further validated by Damron et al. [2] for lesions in the femur as a separate group [9]. In both the original manuscript [9] and the more recent independent validation for femoral lesions [2], the optimal score for impending pathologic fracture predicts that approximately ⅓ of those bones will fracture. By contrast, the total Mirels score identified in this study as optimal for identifying impending pathologic fractures of the humerus is seven or greater. Lowering the threshold for defining a lesion as an
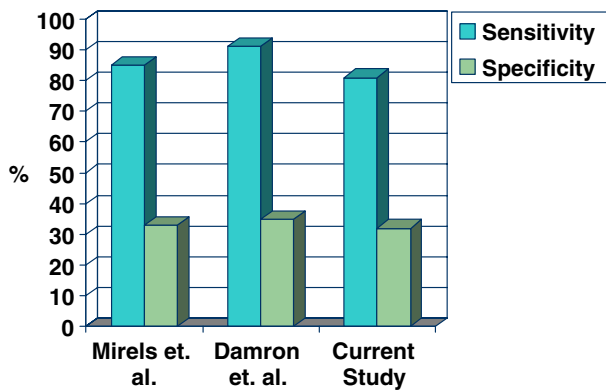
Fig. 2 We compared Mirels' sensitivity and specificity profiles from three studies. Lowering the threshold for defining a lesion as an impending pathologic fracture from nine points in Mirels [9] and Damron et al. [2] to seven for the humerus appears to preserve the same level of sensitivity and specificity that the Mirels rating system displays for other long bones.

impending pathologic fracture to seven for the humerus appears to preserve the same level of sensitivity and specificity that the Mirels rating system displays for other long bones (Fig. 2). A threshold of nine in this study resulted in an unacceptable sensitivity level of 81.4% with progressive enhancement of sensitivity up to 32.1% at a threshold of seven. To the best of our knowledge, this is the first study in the literature to evaluate the Mirels rating system as applied to metastatic lesions specifically in the humerus.

Use of Mirels' rating system has been criticized for its poor specificity [2]. Using definitions of nine or greater for femoral and seven or greater for humeral impending pathologic fractures yields sensitivity of 80% to 90% and specificity of 30% to 35%. Unfortunately, this suggests that using the Mirels rating system may result in unnecessary prophylactic stabilization of lesions identified as impending pathologic fractures. Moreover, because sensitivity is less than 100%, 10% to 20% of impending pathologic fractures may be missed using these definitions. The morbidity associated with performing unnecessary surgery includes exposure of the patient to the risks of surgical and anesthetic complications, postoperative complications as well as potential mismatch of actual surgical recovery time and projected life expectancy. On the other hand, there is published evidence suggesting poorer function when pathologic lesions are allowed to progress to fracture rather than fixing them prophylactically. Flemming and Beals [4] reported a high frequency of inadequate perioperative pain control in patients who fractured pathologically and later underwent definitive stabilization at a nonunion rate of 50%. Over 50% of those patients had poor to fair functional outcome. The avoidance of unnecessary surgery and the prevention of pathologic fracture are critical priorities when aiming to minimize the morbidity of metastatic bone disease.

The Mirels rating system [9] is a valid, reproducible screening tool for identifying impending pathologic fractures of the humerus when used by physicians across experience levels. In our study expanding upon previous use in all long bones and in only femoral bones, it remained reproducible, valid, and more effective than clinical judgment alone when predicting fracture risk and when making treatment decisions regarding the use of prophylactic stabilization in the humerus. Further, as a composite clinical tool made up of four individual components, this weighted scoring system performs better than its individual components if used independently. Among currently available clinical guidelines for long bone fracture prediction in the setting of metastatic disease and myeloma, Mirels' system remains the preferred choice. However, more specific guidelines are needed to more selectively predict fracture risk in patients with metastatic disease to long bones.

Appendix 1. Mirels Scoring Evaluation Study 3/04

Examiner_____

Sample History

Case

A 68-year-old female with no known malignancy presented to the clinic with progressive soreness of her right upper extremity over the past 2 to 3 weeks. She described the pain as "moderate," but worsening with activity. She has not been taking any medication for pain control. Patient has been wearing a Sarmiento brace for the past week until her appointment. Further testing discovered a primary malignancy of infiltrating ductal breast carcinoma with metastases to her liver, lung, and femur. Patient denies other past medical history. The patient desires adequate pain control.
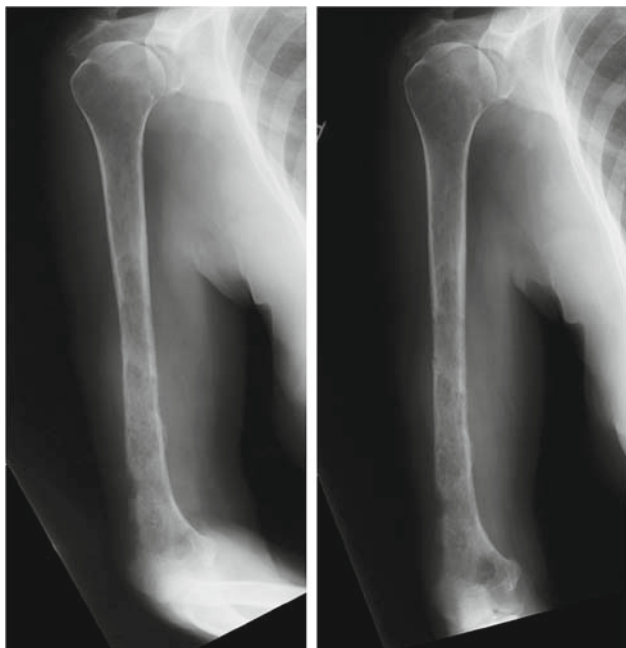
MIRELS SCORING: (circle one)

PAIN	1	2	3
LOCATION	1	2	3
SIZE	1	2	3
BLASTIC/MIXED/LYTIC	1	2	3

Total Score: 1   2   3   4   5   6

7   8   9   10   11   12

DOES THIS PRESENT AN IMPENDING PATHOLOGIC FRACTURE IN YOUR OPINION? (regardless of Mirels score)

Y   N

COMMENTS:

References

1. Coleman RE. Metastatic bone disease: clinical features, pathophysiology, and treatment strategies. *Cancer Treat Rev*. 2001;27:165–176.
2. Damron TA, Morgan H, Prakash D, Grant W, Aronowitz J, Heiner J. Metastatic disease of long bones: critical evaluation of Mirels' Rating System for impending pathologic fractures. *Clin Orthop Relat Res*. 2003;415S:S201–S207.
3. Fidler M. Prophylactic internal fixation of secondary neoplastic deposits in long bones. *BMJ*. 1973;1:341–343.
4. Flemming JE, Beals RK. Pathologic fractures of the humerus. *Clin Orthop Relat Res*. 1986;203:258–260.
5. Green AG. Kappa Statistics for Multiple Raters Using Categorical Classifications. Proceedings of the Twenty-Second Annual SAS Users Group International Conference. Cary NC: SAS Institute; 1997.
6. Harrington KD. Impending pathologic fractures from metastatic malignancy: evaluation and management. *Instr Course Lect*. 1986;35:357–381.
7. Hipp JA, Springfield DS, Hayes WC. Predicting pathologic fracture risk in the management of metastatic bone defects. *Clin Orthop Relat Res*. 1995;312:120–135.
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
9. Mirels H. Metastatic disease in long bones: a proposed scoring system for diagnosing impending pathologic fractures. *Clin Orthop Relat Res*. 1989;249:256–264.
10. Parrish FF, Murray JA. Surgical treatment for secondary neoplastic fractures. *J Bone Joint Surg Am*. 1970;52:665–686.
11. Snell WE, Beals RK. Femoral metastases and fracture from breast cancer. *Surg Gynecol Obstet*. 1964;119:22–24.
12. Zickel RE, Mouradian WH. Intramedullary fixation of pathological fractures and lesions of the subtrochanteric region of the femur. *J Bone Joint Surg Am*. 1976;58:1061–1066.