**Clinical Orthopaedics
and Related Research®**
A Publication of The Association of Bone and Joint Surgeons®

CrossMark

CLINICAL RESEARCH

Do Orthopaedic Oncologists Agree on the Diagnosis and Treatment of Cartilage Tumors of the Appendicular Skeleton?

Tomas Zamora MD, Julio Urrutia MD, Daniel Schweitzer MD,
Pedro Pablo Amenabar MD, Eduardo Botello MD

Abstract

Background   Distinguishing a benign enchondroma from a low-grade chondrosarcoma is a common diagnostic challenge for orthopaedic oncologists. Low interrater agreement has been observed for the diagnosis of cartilaginous neoplasms among radiologists and pathologists, but, to our knowledge, no study has evaluated inter- and intraobserver agreement among orthopaedic oncologists grading these lesions using initial clinical and imaging information. Determining such agreement is important since it reflects the certainty in the diagnosis by orthopaedic oncologists. Agreement also is important as it will guide future treatment and prognosis, considering that there is no gold standard for diagnosis of these lesions.

Questions/Purposes   (1) to determine inter- and intraobserver agreement among a multinational panel of expert orthopaedic oncologists in diagnosing cartilaginous neoplasms based on their assessment of clinical symptoms and imaging at diagnosis. (2) To describe the most important clinical and imaging features that experts use during the initial diagnostic process. (3) To determine interobserver agreement for proposed initial treatment strategies for cartilaginous neoplasms by this panel of evaluators.

Methods   Thirty-nine patients with intramedullary cartilaginous neoplasms of the appendicular skeleton of various histopathologic grades were selected and classified as having benign, low-grade malignant, or intermediate- or high-grade malignant neoplasms by 10 experienced orthopaedic oncologists based on clinical and imaging information. Additionally, they chose the three most important clinical or imaging features for the diagnosis of these neoplasms, and they proposed a treatment strategy for each patient. The Kappa coefficient (κ) was used to determine inter- and intraobserver agreement.

Results   Inter- and intraobserver agreements were only fair to good, κ = 0.44 (95% CI, 0.41–0.48) and κ = 0.62 (95% CI, 0.52–0.72), respectively. The three factors most frequently identified as helpful in making the diagnosis by our panel were cortical involvement in 65% of evaluations (253/390), neoplasm size in 51% (198/390), and pain in 50% (194/390). The interobserver agreement for the proposed initial treatment strategy after diagnosis was poor (κ = 0.21; 95% CI, 0.18–0.24).

Conclusions   This study showed barely fair interobserver and fair to good intraobserver agreement for grading of intramedullary cartilaginous neoplasms by orthopaedic oncologists using initial clinical and imaging findings. These results reflect the insufficient guidance interpreting clinical and imaging features, and the limitations of the systems we use today when making these diagnoses. In the same way, they generate concern for the implications that this may have on different treatment strategies and the future prognosis of our patients. Future studies should build

T. Zamora, J. Urrutia, D. Schweitzer, P. P. Amenabar, E. Botello (✉)
Department of Orthopaedic Surgery, School of Medicine, Pontificia Universidad Catolica de Chile, Diagonal Paraguay 362, 8330077 Santiago, Chile
e-mail: ebotello@med.puc.cl

on these observations and focus on clarifying our criteria of diagnosis so that treatment recommendations are standardized regardless of the treating institution or oncologist. *Level of Evidence*   Level III, diagnostic study.

Introduction

Cartilaginous neoplasms are among the most common tumors of the appendicular skeleton and can involve almost any bone [20]. Distinguishing between a benign enchondroma and a low-grade chondrosarcoma is a common diagnostic challenge for orthopaedic oncologists [9, 21, 31]. Although multiple clinical and radiologic elements have been shown to aid in the diagnostic process [1, 9, 12, 21], accurate differentiation between benign and malignant cartilaginous neoplasms can be difficult, even for trained specialists. Furthermore, no specific gold standard has been developed to resolve this diagnostic challenge. Advanced imaging has been of great aid performing these diagnoses, but radiography, CT, and MRI have limitations [2, 4]. In addition, even pathologists have difficulties establishing these diagnoses [27], with a lower than an desirable agreement between experts.

Even when patients are seen for consultation by a multidisciplinary team, the initial decision whether to observe a patient or perform a biopsy or other more-invasive procedures frequently is made by the orthopaedic oncologist based on his or her global evaluation of clinical and imaging elements. An interrater agreement study showed low agreement among experienced pathologists and radiologists for grading cartilaginous lesions [27]; however, to the best of our knowledge, no agreement assessments have been reported for orthopaedic oncologists. Determining agreement among orthopaedic oncologists diagnosing cartilaginous neoplasms is important because it reflects the level of certainty that specialists have when assessing these patients; disagreement by clinicians in this field can lead to an entirely different course of action for a patient, even when consulting with highly experienced specialists. This ability to distinguish benign from malignant cartilaginous tumors could affect treatment times, the number of procedures performed, and even patients' survival. Therefore, understanding the agreement of orthopaedic oncologists and making a specific diagnosis is a critical step, and is a fundamental element if we try to elaborate a more efficient diagnostic and treatment algorithm for these types of tumors.

Consequently, we performed this study (1) to determine inter- and intraobserver agreement among a multinational panel of expert orthopaedic oncologists in diagnosing cartilaginous neoplasms based on their assessment of clinical symptoms and imaging at diagnosis, (2) to describe the most important clinical and imaging features that experts use during the initial diagnostic process, and (3) to determine interobserver agreement for proposed initial treatment strategies for cartilaginous neoplasms by this panel of evaluators.

Patients and Methods

Study Design and Setting

This study was approved by the ethics review board of our institution. We collected and analyzed the data for 39 patients with an intramedullary cartilaginous neoplasm of the appendicular skeleton (proximal to the metacarpals and metatarsals) from a large database with 550 patients who were treated by one surgeon (EB) between 2005 and 2012.

Participants/Study Subjects

A similar proportion of patients with tumors of three histopathologic grades (benign, low-grade malignant, and intermediate- or high-grade malignant) of the appendicular skeleton including the pelvis and scapula were included in the study, and were selected for their heterogeneity and representability. In patients with neoplasms with a clear benign aspect and no histologic study, the inclusion criterion was the observance of the same benign character (without changes in appearance, size on images, or any other change on clinical condition, pain included) for at least 3 years after the original diagnosis. For all other patients, the inclusion diagnosis and criteria were a confirmed histopathologic diagnosis of enchondroma or chondrosarcoma by a fellowship-trained musculoskeletal pathologist based on characteristic features such as the presence of hyaline cartilage, cellularity, nuclear pleomorphism, cell atypia, invasion of adjacent structures, foci of necrosis, and calcifications, among others. Complete clinical data with a detailed physical examination and available imaging studies were necessary for inclusion. We excluded patients with tumors that were located in the metacarpals, metatarsals, phalanges, and spine. Additionally, we excluded patients with a final histopathologic diagnosis or who had a clinical scenario indicative of an osteochondroma, chondroblastoma, or chondromyxoid fibroma.

No pathologists or radiologists were directly involved in our study; the senior author (EB) used the following criteria only for patient selection: a diagnosis made by pathologist, or a clinical and radiologic followup for more than 3 years for patients with a tumor with benign appearance but without a histologic diagnosis. Of the 39 patients, 14 had benign neoplasms, 13 had low-grade

Table 1. Patients' clinical information and available images

Patient number	Gender	Age (years)	Diagnosis	Affected bone	Image available
1	Male	39	E	Distal femur	Radiographs and MRI
2	Female	59	LGC	Proximal humerus	Radiographs and MRI
3	Female	43	LGC	Proximal humerus	CT and MRI
4	Female	54	LGC	Proximal humerus	CT and MRI
5	Female	68	E	Distal tibia	Radiographs and MRI
6	Female	25	LGC	Scapula	Radiographs, CT, and MRI
7	Male	77	IHGC	Proximal femur	Radiographs and MRI
8	Male	51	E	Proximal humerus	Radiographs and CT
9	Male	57	LGC	Distal femur	Radiographs
10	Female	60	IHGC	Pelvis	Radiographs and MRI
11	Male	63	E	Distal femur	Radiographs and MRI
12	Male	55	E	Distal femur	Radiographs and CT
13	Male	52	IHGC	Pelvis	Radiographs and MRI
14	Male	69	IHGC	Pelvis	Radiographs and MRI
15	Male	66	E	Distal femur	Radiographs and MRI
16	Female	50	E	Proximal humerus	CT
17	Female	55	E	Distal femur	Radiographs and CT
18	Female	45	E	Proximal humerus	Radiographs and MRI
19	Female	26	E	Proximal femur	Radiographs, CT, and MRI
20	Female	27	E	Proximal humerus	CT and MRI
21	Male	48	IHGC	Proximal humerus	Radiographs, CT, and MRI
22	Female	63	LGC	Fibula	Radiographs and CT
23	Female	45	IHGC	Pelvis	Radiographs, CT, and MRI
24	Female	84	IHGC	Proximal femur	Radiographs, CT, and MRI
25	Female	65	E	Distal femur	Radiographs and CT
26	Male	56	LGC	Distal femur	Radiographs and MRI
27	Male	48	E	Proximal humerus	Radiographs
28	Male	27	IHGC	Pelvis	Radiographs, CT, and MRI
29	Male	48	LGC	Distal femur	Radiographs and MRI
30	Male	58	IHGC	Fibula	Radiographs and MRI
31	Female	51	LGC	Proximal humerus	Radiographs and MRI
32	Female	48	E	Distal femur	Radiographs and MRI
33	Female	34	IHGC	Pelvis	Radiographs and MRI
34	Female	85	IHGC	Proximal femur	Radiographs
35	Female	55	LGC	Proximal femur	Radiographs, CT, and MRI
36	Male	42	LGC	Pelvis	Radiographs, CT, and MRI
37	Female	65	IHGC	Proximal femur	Radiographs and MRI
38	Female	70	LGC	Pelvis	Radiographs, CT, and MRI
39	Female	28	LGC	Proximal humerus	Radiographs and MRI

E = enchondroma; LGC = low-grade chondrosarcoma; IHGC = intermediate/high-grade chondrosarcoma.

malignant neoplasms, and 12 had intermediate- or high-grade malignant neoplasms (Table 1).

Description of Experiment, Treatment, or Surgery

The senior author (EB), who did not act as an evaluator, gathered all the data that were necessary to perform the assessment. The data included clinical information (age, gender, presence or absence of pain and its relation to rest or activity, medication requirements, the form of presentation, time of progression, and physical examination findings), radiographs and other available images (MRI or CT). Clinical information needed a detailed description of the patient's condition to ensure proper analysis of clinical features. All patients had at least biplanar radiographs or a

CT scan, and most also had an MRI, presented digitally with the most-representative slices. The information from each of the 39 patients was sent to 10 expert orthopaedic oncologists from different areas in Latin America, Spain, and Italy. All of the experts had more than 5 years of dedicated work on the field, and they are leading members of an orthopaedic oncology reference center in their countries. They were not involved in the patients' care, and they did not know the patients' identities, their definitive diagnosis, or their treatment. Additionally, they were unaware of distribution of the patients' diagnosis. The 10 evaluators were asked to grade each lesion as either (1) benign, (2) low-grade malignant, or (3) intermediate- or high-grade malignant considering all clinical and imaging features available. Assessors had to consider some selected clinical and imaging features (Table 2), and weight their value according to their clinical experience and knowledge to make the diagnosis. Additionally, the evaluators rated the three top clinical or radiologic elements that guided their diagnostic hypothesis for each patient using their judgment (patient age, the amount of cortical involvement, neoplasm size, pain, tumor location, presence or absence of calcifications). Finally, the evaluators were asked to choose one of the following four initial treatment plans for each patient based on the available clinical and radiologic information: followup with sequential clinical assessment and radiographic evaluation; percutaneous or open incisional biopsy; curettage with or without adjuvant treatment; or wide or radical surgical resection. Interobserver agreement was determined by comparing the initial responses of the 10 evaluators. Intraobserver agreement was determined by comparing two assessments of the same patient by the same evaluator; the two assessments were separated by a 3-month interval, and the patients were presented with the same clinical scenario and images in a random sequence to avoid recall bias as reported in other studies [17, 28, 29].

Statistical Analysis and Study Size

R software (R Project for Statistical Computing, Vienna, Austria) was used to determine the necessary sample size

Table 2. Main clinical and imaging features evaluated by experts

Patient's age

Clinical presentation

Presence of pain and its relation to rest and activity

Tumor localization and growth

Amount of cortical involvement observed on images

Neoplasm size

Presence or absence of calcifications on images

to achieve statistical significance. To our knowledge, this is the first agreement study of orthopaedic oncologists for the differential diagnosis and treatment of cartilaginous neoplasms. Assuming similar agreement to those observed for other specialties [27], a confidence interval approach for sample-size estimation with multiple raters, as reported by Rotondi and Donner [25], was used for the interobserver agreement studies. With an expectation of fair to good reliability, we estimated a 95% CI with a lower limit of 0.3 and an upper limit of 0.6, and this resulted in an estimated required sample size of 35 patients.

SPSS software, Version 20.0 (IBM Corp, Armonk, NY, USA) was used for statistical analysis. Fleiss' kappa (κ) coefficient [11] was used to determine the interobserver agreement for the orthopaedic oncologists' diagnostic hypotheses and proposed treatment strategies. The κ coefficient is the most-commonly used agreement statistic in medical studies as it indicates the magnitude of exact agreement between different evaluators with correction by chance. Fleiss' κ coefficient was designed for assessment of agreement between multiple raters and a categorical rating, in contrast to the most-used form of κ coefficient, Cohen's κ, which is designed for only two evaluators. Levels of agreement were determined as proposed by Fleiss: κ values less than 0.40, 0.40 to 0.75, and greater than 0.75, indicated poor, fair to good, and excellent agreement, respectively [11]. The κ values are presented with 95% CIs.

Considering that the κ values are affected by the prevalence of the phenomenon evaluated, we decided to use a similar proportion of patients with tumors of three categories; thus, κ values for each category would truly reflect agreement and should not be influenced by the prevalence of each lesion.

Finally, the diagnosis proposed by all 10 evaluators was compared with the diagnosis of inclusion of each patient; such comparison was reported as a proportion and with κ values for agreement analysis.

Results

Full interobserver agreement among all 10 evaluators was achieved in only five of 39 patients (13%), with $\kappa = 0.44$ (95% CI, 0.41–0.48), indicating fair to good agreement (Table 3). Only four of five patients had full agreement that their neoplasm was considered benign by all 10 evaluators (Fig. 1), and in one it was considered intermediate- or high-grade malignant (Fig. 2). The κ values for the interobserver agreement for each diagnostic hypothesis were as follows: 0.51 (95% CI, 0.47–0.56) for enchondromas, 0.21 (95% CI, 0.16–0.26) for low-grade chondrosarcomas, and 0.6 (95% CI, 0.55–0.64) for intermediate- or high-grade chondrosarcomas.

Table 3. Evaluators' responses for each patient

Patient number	Evaluator 1		Evaluator 2		Evaluator 3		Evaluator 4		Evaluator 5		Evaluator 6		Evaluator 7		Evaluator 8		Evaluator 9		Evaluator 10	
	Diag	Treat	Diag	Treat	Diag	Treat	Diag	Treat	Diag	Treat	Diag	Treat	Diag	Treat	Diag	Treat	Diag	Treat	Diag	Treat
1	E	O	E	O	E	O	E	O	E	O	E	O	E	C	E	O	E	B	E	O
2	LGC	B	IHGC	B	LGC	C	LGC	B	E	O	E	O	LGC	C	E	O	LGC	B	LGC	R
3	LGC	B	LGC	B	LGC	C	LGC	B	LGC	C	LGC	B	E	C	LGC	R	LGC	B	LGC	R
4	E	O	LGC	B	E	C	LGC	B	LGC	C	LGC	B	E	C	E	O	E	B	LGC	R
5	E	O	E	O	E	O	E	O	E	O	E	O	E	C	E	O	E	O	E	O
6	IHGC	C	E	O	LGC	B	E	O	LGC	C	LGC	R	E	C	LGC	R	E	B	LGC	R
7	IHGC	R	IHGC	R	IHGC	R	IHGC	B	IHGC	B	IHGC	R	IHGC	R	IHGC	R	IHGC	B	IHGC	R
8	E	O	E	O	E	O	E	O	E	O	E	O	E	C	E	O	E	O	E	O
9	E	O	E	O	E	O	LGC	B	E	O	LGC	B	LGC	C	E	O	E	B	LGC	C
10	LGC	B	IHGC	B	IHGC	B	LGC	B	IHGC	B	IHGC	R	IHGC	R	LGC	R	IHGC	B	LGC	R
11	E	O	E	O	E	O	E	O	E	O	E	O	E	C	E	O	E	B	E	O
12	E	O	IHGC	B	E	C	E	O	E	O	E	O	E	C	E	O	LGC	B	E	O
13	IHGC	R	IHGC	R	IHGC	R	IHGC	B	IHGC	B	IHGC	R	IHGC	B	LGC	R	IHGC	B	IHGC	R
14	IHGC	R	IHGC	R	IHGC	R	IHGC	B	IHGC	B	IHGC	R	IHGC	B	LGC	R	IHGC	B	IHGC	R
15	E	O	LGC	B	E	O	E	O	E	O	LGC	O	LGC	C	E	O	E	B	E	O
16	LGC	B	IHGC	R	E	B	LGC	B	LGC	B	LGC	B	E	B	E	O	LGC	B	E	O
17	E	O	LGC	B	E	O	E	O	E	O	E	O	E	C	E	O	E	B	E	O
18	E	O	LGC	B	E	O	LGC	B	E	O	E	O	E	C	E	O	E	B	E	O
19	E	O	LGC	B	E	C	E	O	LGC	B	LGC	B	E	C	E	O	E	B	E	C
20	LGC	C	LGC	B	E	C	LGC	B	E	O	E	O	E	B	E	O	LGC	B	LGC	C
21	IHGC	B	IHGC	R	E	C	LGC	B	IHGC	B	LGC	B	E	B	E	O	LGC	B	E	O
22	LGC	R	E	B	E	C	LGC	R	LGC	R	LGC	B	LGC	C	LGC	O	IHGC	B	E	O
23	IHGC	R	IHGC	B	IHGC	R	LGC	B	IHGC	B	IHGC	R	IHGC	B	IHGC	R	IHGC	B	IHGC	R
24	IHGC	R	IHGC	R	LGC	R	IHGC	B	IHGC	R	IHGC	R	IHGC	B	LGC	R	E	B	LGC	O
25	LGC	B	LGC	B	E	O	E	O	E	O	LGC	B	E	B	E	O	LGC	B	E	O
26	LGC	B	LGC	B	LGC	C	LGC	B	LGC	C	LGC	B	LGC	B	E	O	E	B	LGC	C
27	E	O	E	O	E	C	E	O	E	O	E	O	LGC	B	E	O	IHGC	B	E	O
28	IHGC	R	IHGC	R	IHGC	R	LGC	B	IHGC	B	IHGC	R	IHGC	B	IHGC	R	LGC	B	IHGC	R
29	E	O	E	B	E	C	E	O	E	O	LGC	B	E	B	E	O	IHGC	B	E	O
30	LGC	R	IHGC	R	LGC	R	LGC	R	LGC	R	IHGC	R	IHGC	B	LGC	R	LGC	B	E	R
31	E	O	LGC	B	E	O	LGC	B	LGC	C	E	B	E	B	E	O	E	O	E	O
32	E	O	E	B	E	O	E	O	E	O	LGC	B	E	B	E	O	IHGC	B	LGC	C
33	IHGC	R	IHGC	R	LGC	R	E	O	IHGC	B	IHGC	R	IHGC	B	IHGC	R	IHGC	B	IHGC	R
34	IHGC	R	IHGC	R	IHGC	R	LGC	B	IHGC	R	IHGC	R	IHGC	R	LGC	R	IHGC	B	IHGC	R

Table 3. continued

Patient number	Evaluator 1		Evaluator 2		Evaluator 3		Evaluator 4		Evaluator 5		Evaluator 6		Evaluator 7		Evaluator 8		Evaluator 9		Evaluator 10	
	Diag	Treat	Diag	Treat	Diag	Treat	Diag	Treat	Diag	Treat	Diag	Treat	Diag	Treat	Diag	Treat	Diag	Treat	Diag	Treat
35	IHGC	R	LGC	R	LGC	R	E	B	IHGC	B	IHGC	R	IHGC	B	LGC	R	LGC	B	LGC	R
36	IHGC	R	IHGC	R	IHGC	R	E	B	IHGC	B	IHGC	R	IHGC	B	IHGC	R	IHGC	B	IHGC	R
37	LGC	B	IHGC	R	LGC	R	IHGC	B	IHGC	B	IHGC	R	IHGC	B	LGC	R	IHGC	B	IHGC	R
38	LGC	B	IHGC	B	LGC	R	LGC	B	IHGC	B	LGC	R	LGC	B	LGC	R	IHGC	B	LGC	R
39	E	O	IHGC	B	LGC	R	LGC	B	IHGC	B	LGC	B	LGC	B	LGC	R	LGC	B	LGC	R

E = enchondroma; LGC = low-grade chondrosarcoma; IHGC= intermediate/high-grade chondrosarcoma; Diag = diagnosis; Treat = initial managment; O = observe; B = biopsy; C = curettage; R = resection.

The level of agreement was considered fair to good for benign and intermediate- or high-grade malignant neoplasms but poor for low-grade malignant neoplasms.

The imaging modalities available varied among patients: radiographs, CT, and MRI were available for 33 of 39, 18 of 39, and 30 of 39 patients, respectively. However, all patients had at least biplanar radiographs or a CT scan, and most also had MR images. The analysis of interobserver agreement for subgroups showed a κ = 0.45 (95% CI, 0.42–0.49) for the 30 patients with an additional MRI, which is similar to the κ value obtained for the whole group. For the repeat evaluation 12 weeks after the first assessment, we observed 75% full intraobserver agreement for distinguishing between benign, low-grade, and inter-mediate- or high-grade malignant cartilaginous neoplasms. The κ value was 0.62 (95% CI, 0.52–0.72), which indicates fair to good reliability.

If we excluded pelvic lesions (analyzing only the 31 patients with long-bone tumors), the interobserver agreement showed a κ value of 0.37 (95% CI, 0.33–0.41), which indicates poor reliability.

When the orthopaedic oncologists were asked to list the three most-important clinical or imaging features that influenced their diagnostic hypotheses, cortical involvement, neoplasm size, and pain were chosen in 65% (253/390), 51% (198/390), and 50% (194/390) of the evaluations, respectively. Tumor location, patient's age, and the presence or absence of calcifications were chosen less frequently (35%, 29%, and 14% of evaluations, respectively) and therefore were considered to be less useful for the experts' opinions.

The interobserver analysis of the proposed treatment strategies showed only poor agreement, with κ = 0.21 (95% CI, 0.18–0.24). Different proportions of recommended treatments were observed for different diagnoses proposed by evaluators. For lesions diagnosed as enchondroma (162 evaluations), the evaluators suggested observation and radiologic followup, open or percutaneous incisional biopsy, curettage with or without adjuvant treatment, or wide surgical resection in 70% (115/162), 17% (28/162), 11% (18/162), and 1% (one of 162) of patients, respectively. For tumors diagnosed as low-grade chondrosarcoma (117 evaluations), the assessors proposed biopsy, wide surgical resection (if possible), and curettage with or without adjuvant treatment in 54% (63/117), 32% (37/117), and 14% (17/117) of patients, respectively. Finally, for lesions diagnosed as intermediate- or high-grade chondrosarcoma (111 evaluations), wide resection, open or percutaneous biopsy, and curettage with or without adjuvant treatment were proposed in 54% (60/111), 45% (50/111), and 1% (one of 111) of patients, respectively. No evaluator proposed observation or followup for lesions considered as malignant neoplasm.

Fig. 1A–B A 39-year-old man without any previous pain obtained plain radiographs after a knee contusion while playing soccer. He had no pain at the distal femur or any other relevant findings on physical examination. (**A**) An AP radiograph of his distal femur shows a central lesion with a cartilaginous matrix and calcifications. (**B**) A coronal T1 MR image shows a 2 × 4-cm cartilaginous tumor without any cortical compromise or soft tissue mass. This was interpreted as benign by all evaluators.

Fig. 2A–B A 77-year-old man had a history of left groin pain for 3 months. His pain progressively increased during the last weeks, even at rest, with substantial volume enlargement of the proximal left thigh. The physical examination disclosed pain with left hip movements and a palpable soft tissue mass on the anteromedial thigh. (**A**) An AP radiograph of his left hip shows a large lytic lesion with central calcifications and ill-defined borders on the proximal femur. (**B**) A coronal T2 MR image shows the same cartilaginous tumor with extensive compromise until the subtrochanteric area and medial soft tissue mass. This was interpreted as intermediate- or high-grade malignant by all evaluators.
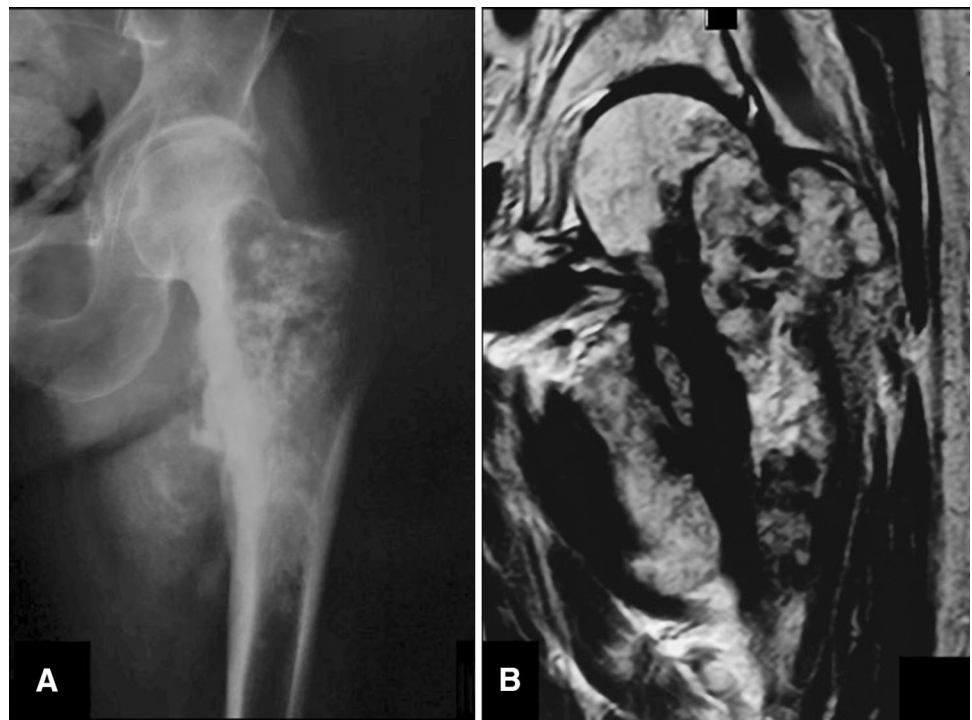
Our experts agreed with patients' initial inclusion diagnoses in 70% of evaluations (273/390). This result revealed a fair to good agreement with a kappa value of 0.54 (95% CI, 0.47–0.61).

Discussion

Differentiating between an enchondroma and a chondrosarcoma, especially low-grade chondrosarcoma,

remains a diagnostic challenge, even for experienced musculoskeletal oncologists [9]. Although accurate diagnosis generally involves an evaluation by a multidisciplinary team, the orthopaedic oncologist usually has to integrate most of the information and make the initial decision regarding how to proceed. It is clear from previous studies [6, 12, 27] that radiologists and even pathologists frequently disagree on a definitive diagnosis when facing cartilaginous tumors; this limitation means that we are left without any true gold standard to diagnose these lesions. Therefore, an inter- and intraobserver agreement study using clinicians as evaluators diagnosing these lesions is important because the orthopaedic oncologists' initial diagnosis will define the initial management implemented, and their disagreement can lead to an entirely different course of action and possibly affect the patients' outcomes. Thus, to identify how accurate the clinicians are with their diagnosis and initial treatment strategies could be an important step in determining the reasons for the mixed results that we have had with time with this type of tumor, especially when considering oncologic survival [22]. Our results showed only fair interobserver agreement for diagnosis and grading of cartilaginous neoplasms and poor agreement for proposed patient management by orthopaedic oncologists. These findings show a worrisome scenario since they reflect that a patient may receive a completely different diagnosis and initial management even by trained expert orthopaedic oncologists. Furthermore, these results are similar to those of other studies analyzing agreement between specialized and experienced pathologists and radiologists [6, 12, 27].

This study has several limitations. First, not all patients had a confirmed histopathologic diagnosis. The recommendation of most orthopaedic oncologists for intramedullary tumors presumed to have a benign behavior on clinical and radiologic analysis is followup without any surgical intervention or biopsy; moreover, it would have been unethical to perform an additional procedure solely for research purposes. However, the Guidelines for Reporting Reliability and Agreement Studies [27] specifically emphasized the need for representativeness and external validity for reliability and agreement studies. We intended to achieve representativeness and external validity by including the entire spectrum of patients encountered on a regular basis by an orthopaedic oncologist, from patients with clearly benign enchondromas, which are only carefully followed, to patients with high-grade chondrosarcomas, which usually are treated with surgical resection. Conversely, it is clear from previous studies that not even pathologists or radiologists always agree on a definitive diagnosis when facing cartilaginous tumors. Therefore, determining inter- and intraobserver agreement study is important because the

results do not depend on the initial definition of a patient's diagnosis, but mainly on the reliability of evaluators between each other and the reproducibility among themselves. However, it could be considered that we used only the clinician's diagnosis and treatment suggestion in our study, whereas the best practice for oncologic lesions is a multidisciplinary team evaluation (by orthopaedic oncologists, radiologists, and pathologists). We chose this approach in our study because we specifically wanted to evaluate how orthopaedic surgeons approach these lesions and whether they agree with their diagnosis and their initial action taken. Our results showed that even experienced specialists frequently disagree on this scenario. Future studies should include a more-complex and multidisciplinary approach.

In this study, we included three possible alternatives for diagnosis: (1) benign, (2) low-grade malignant, and (3) intermediate- or high-grade malignant cartilaginous tumors. Differentiating a low-grade chondrosarcoma from a benign tumor is a diagnostic challenge and is reflected here by our results. In the same way, we included four initial actions proposed by the evaluators: (1) to observe, (2) to perform a percutaneous or incisional biopsy for further diagnosis, or to directly treat either with (3) curettage or a (4) wide resection. The inclusion of an intermediate alternative in diagnosis (low-grade chondrosarcoma) and on treatment strategies (to perform a biopsy) could influence our lower agreement, but it represents the real options that an orthopaedic oncologist faces when diagnosing and treating these patients. Future studies should address agreement between benign tumors and higher-grade chondrosarcomas separately and also specific final treatment options after a complete clinical evaluation has been done.

Another limitation of our study is that not all patients had the same form of imaging performed. This again represents the real clinical scenario and it is part of the limitations of collecting patients on a retrospective basis. Even when most of our patients had radiographs, CT, and MRI performed for their evaluation, it raises the question whether the agreement was affected by this difference, and potentially better outcomes could be achieved by a protocol driven by image acquisition. To address this limitation, a subgroup analysis of patients who had an additional form of evaluation (MRI) was performed; such evaluation did not reveal differences in the agreement obtained for those particular patients. Although our results are similar to those reported previously [27], this particular issue should be evaluated in future studies using a protocol driven by image acquisition.

When considering results of any reliability study, emphasis should be put on the specific clinical scenario that is being discussed and the relevance of the agreement (or disagreement) that is being reported. On clinical matters, adjectives frequently used to give sense to

abstract values such as poor, fair to good, or excellent, fail to clearly represent what the real consequences of an incorrect diagnosis or a different treatment plan are. This is clear when comparing a κ value representing fair to good agreement, and complete agreement was achieved in only five of 39 patients. That is a large limitation when analyzing this type of study. Finally, there is no validated form of assessing intraobserver agreement on this scenario. Most studies allow 3 to 12 weeks between evaluations to diminish possible memory bias. Our evaluators did the second round of ratings 12 to 24 weeks after the initial one (depending on availability of each evaluator, the survey was sent at Week 12). Even when there is no validated form of doing this, we believe that intraobserver agreement analysis is important because it represents the variability that a diagnosis can have for the same evaluator at a different time, and 12 to 24 weeks is reasonable for doing so.

This study showed barely fair interobserver agreement for the differential diagnosis of benign, low-grade, or intermediate- or high-grade malignant cartilaginous neoplasms, with a κ value of 0.44, but only 13% complete agreement among evaluators. Analysis of these results should be made with caution. Adding more evaluators allows better external validity, but with more experts included in the study, it is more difficult to have complete agreement by all; therefore, the value of the κ statistics should be considered rather than the percentage of agreement. Nevertheless, only five of 39 patients with full agreement on their diagnosis is extremely low. Interobserver evaluations have shown variable agreement for the diagnosis of these types of lesions by other specialists, with κ values of 0.19 to 0.36 for experienced radiologists [2, 12, 27] and 0.44 to 0.78 for pathologists [6, 27]. The benign intramedullary variant of this neoplasm class, enchondroma, is usually asymptomatic and can be treated with observation alone to rule out progression [14]. However, chondrosarcomas range in their grades of malignancy and their capacity to metastasize [13, 18, 19]. Therefore, patients with chondrosarcomas usually undergo surgery ranging from intralesional curettage with or without local adjuvants (such as phenol, ethanol or cement, among others) [5, 30] to wide or radical resection, depending on the histologic grade of the tumor [7, 10, 15, 23, 24, 26].

As expected, greater interobserver diagnostic agreement was observed for clearly benign and highly malignant cartilaginous neoplasms than for low-grade malignant cartilaginous neoplasms. This last group has been
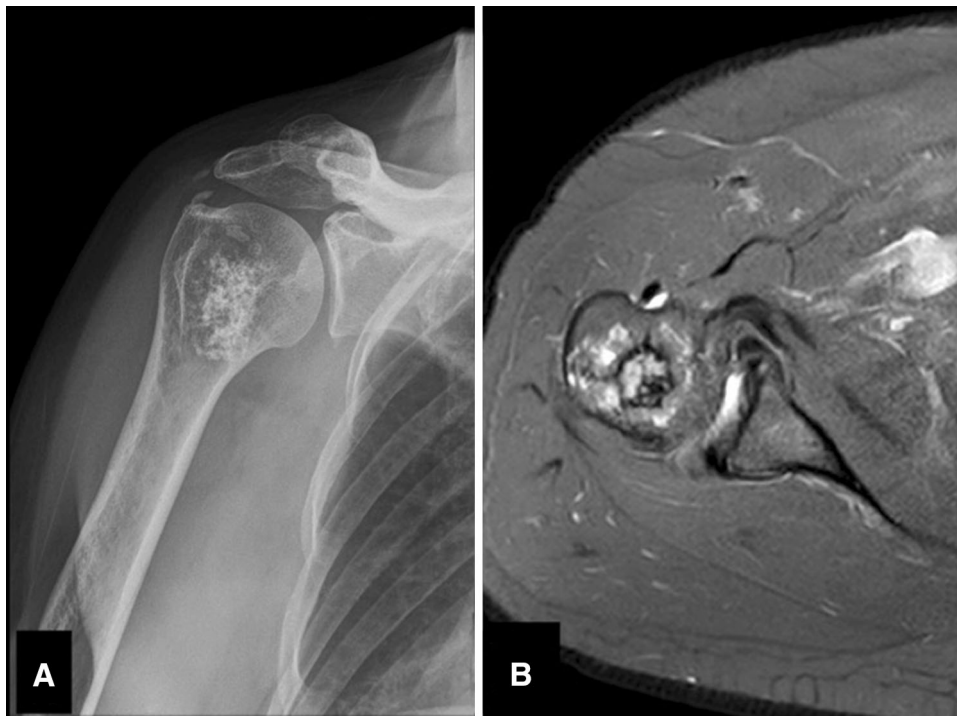

Fig. 3A–B A 51-year-old woman had a history of hypertension and persistent right shoulder pain that worsened at night, with partial response to medication. Her physical examination showed no signs of shoulder impingement and normal active and passive ROM with no modification in pain. (**A**) An AP radiograph of her right shoulder shows a 3 × 4-cm central lesion of the proximal humerus with calcifications and a cartilaginous matrix. (**B**) An axial proton density-weighted MR image shows a cartilaginous tumor of the proximal humerus with a minimal lateral cortical compromise with no soft tissue mass. The final histopathologic diagnosis was a low-grade chondrosarcoma. Only four of 10 evaluators chose the correct diagnosis and poor agreement was observed among them.

described as a unique diagnostic challenge for orthopaedic oncologists [12, 31] (Fig. 3), and this assumption was supported by the finding of lower agreement in the diagnosis of low-grade malignant cartilaginous neoplasms in our study. This result is probably attributable to the limited discriminating power of many of the commonly used imaging parameters, such as cortical compromise, periosteal reaction, tumor diameter, and soft tissue extension [2, 3, 21, 31].

The subgroup analysis according to available image modality did not reveal a difference in diagnosis between the entire group of patients versus those who had multiple image modalities available including plain radiographs, CT, and MRI. These data agree with those of the SLICED study group [27]. Other studies have shown that the ability to diagnose cartilaginous neoplasms is improved by using a combination of imaging modalities, especially the combination of radiographs and MRI, even when increased false positive and false negative findings were observed for the latter [2, 4]. Crim et al. [2] found low interobserver agreement for evaluation of individual imaging criteria and for expert diagnosis based on the results of a single image modality, supporting the need for a multimodality imaging approach.

In this study, the evaluators considered that cortical compromise followed by neoplasm size and the presence or absence of pain were the most useful clinical and radiologic parameters for diagnosis. This finding agrees with those of Murphey et al. [21], who showed that for 187 cartilaginous neoplasms, those same parameters were the most important for distinguishing between enchondroma and low-grade chondrosarcoma. Conversely, patient age, tumor location, and other variables did not influence the diagnosis. Studies by Ferrer-Santacreu et al. [8, 9] showed that pain on palpation, cortical involvement, and bone scan uptake were important factors in the diagnosis of low-grade chondrosarcoma. However, more prospective analyses are needed to validate the actual effect of these variables in the diagnostic algorithm.

Finally, the interobserver agreement for proposed treatment for each lesion was only shown to be poor. This low agreement may be related to the different diagnoses made by the surgeons, but it also may reflect disagreement regarding how to treat a specific tumor according to histologic grade. Nevertheless, most evaluators chose to observe patients with a suspected benign tumor if they were asymptomatic and to perform wide surgical resection or incisional biopsy to confirm the initial diagnosis in neoplasms believed to be intermediate- to high-grade chondrosarcomas, revealing an important consistency between different ways of treating a specific type of tumor. Patients with an initial diagnosis of low-grade chondrosarcoma present the greatest treatment challenge and are probably the ones treated the most

heterogeneously. However, multiple studies recommend curettage with local adjuvant treatments as a harmless way of treating these low-grade malignant tumors, with a low recurrence rate [8, 26, 30].

We believe our results should be followed by future studies including not only clinicians, pathologists, or radiologists, but also a multispecialty team evaluating each patient. Additionally, future studies should be able to weight and establish risk parameters for different features considered in the evaluation, adding relevance to less frequently mentioned elements such as endosteal scalloping, edema pattern on MRI, and the presence of a soft tissue mass. Finally, the inclusion of other imaging modalities such as bone scintigraphy or positron emission tomography-CT [16] should be tested in future evaluations. Meanwhile, a standardized diagnostic algorithm including the most-important clinical features, with an image-driven acquisition protocol and analysis, is key to improving agreement in the diagnosis of cartilaginous neoplasms; such an approach might ensure that patients receive the most appropriate treatment for their condition regardless of the oncologist or institution of treatment.

Achieving an accurate diagnosis of intramedullary cartilaginous neoplasms is challenging, even for experienced orthopaedic oncologists. This study showed only fair to good inter- and intraobserver agreement for grading intramedullary cartilaginous neoplasms based on initial clinical and imaging information. Even more, our results showed poor agreement on the first treatment options based on these features. These results are, without doubt, less than ideal and represent a dilemma if a patient receives widely disparate opinions by different physicians and completely different treatment options. They reflect the lack of guidance regarding how to interpret clinical and imaging features and the limitations of the systems that we use today. A proper diagnosis is central for achieving the best-possible results in terms of survival, complications, and quality of life for patients with cartilaginous neoplasms, so this information emphasizes the need for better diagnostic tools and standardized algorithms to assist clinicians in determining an accurate diagnosis and the most appropriate treatment.

MD (Orthopaedic Oncology Service, Santa Casa de São Paulo, Brasil); Pietro Ruggieri MD (Orthopaedic Oncology Unit, University of Padua. Padua, Italy); and Roberto Reggiani MD (Orthopedic Surgery Service, Universidade Federal de Uberlândia, Uberlândia, Brasil).

References

1. Choi BB, Jee WH, Sunwoo HJ, Cho JH, Kim JY, Chun KA, Hong SJ, Chung HW, Sung MS, Lee YS, Chung YG. MR differentiation of low-grade chondrosarcoma from enchondroma. *Clin Imaging.* 2013;37:542–547.

2. Crim J, Schmidt R, Layfield L, Hanrahan C, Manaster BJ. Can imaging criteria distinguish enchondroma from grade 1 chondrosarcoma? *Eur J Radiol.* 2015;84:2222–2230.

3. De Beuckeleer LH, De Schepper AM, Ramon F. Magnetic resonance imaging of cartilaginous tumors: is it useful or necessary? *Skeletal Radiol.* 1996;25:137–141.

4. De Beuckeleer LH, De Schepper AM, Ramon F, Somville J. Magnetic resonance imaging of cartilaginous tumors: a retrospective study of 79 patients. *Eur J Radiol.* 1995;21:34–40.

5. Di Giorgio L, Touloupakis G, Vitullo F, Sodano L, Mastantuono M, Villani C. Intralesional curettage, with phenol and cement as adjuvants, for low-grade intramedullary chondrosarcoma of the long bones. *Acta Orthop Belg.* 2011;77:666–669.

6. Eefting D, Schrage YM, Geirnaerdt MJ, Le Cessie S, Taminiau AH, Bovee JV, Hogendoorn PC, EuroBoNeT consortium. Assessment of interobserver variability and histologic parameters to improve reliability in classification and grading of central cartilaginous tumors. *Am J Surg Pathol.* J2009;33:50–57.

7. Eriksson AI, Schiller A, Mankin HJ. The management of chondrosarcoma of bone. *Clin Orthop Relat Res.* 1980;153:44–66.

8. Ferrer-Santacreu EM, Ortiz-Cruz EJ, Diaz-Almiron M, Pozo Kreilinger JJ. Enchondroma versus chondrosarcoma in long bones of appendicular skeleton: clinical and radiological criteria: a follow-up. *J Oncol.* 2016;2016:8262079.

9. Ferrer-Santacreu EM, Ortiz-Cruz EJ, Gonzalez-Lopez JM, Perez Fernandez E. Enchondroma versus low-grade chondrosarcoma in appendicular skeleton: clinical and radiological criteria. *J Oncol.* 2012;2012:437958.

10. Fiorenza F, Abudu A, Grimer RJ, Carter SR, Tillman RM, Ayoub K, Mangham DC, Davies AM. Risk factors for survival and local control in chondrosarcoma of bone. *J Bone Joint Surg Br.* 2002;84:93–99.

11. Fleiss JL. *Statistical Methods for Rates and Proportions.* 2nd ed. New York, NY: John Wiley; 1981.

12. Geirnaerdt MJ, Hermans J, Bloem JL, Kroon HM, Pope TL, Taminiau AH, Hogendoorn PC. Usefulness of radiography in differentiating enchondroma from central grade 1 chondrosarcoma. *AJR Am J Roentgenol.* 1997;169:1097–1104.

13. Gitelis S, Bertoni F, Picci P, Campanacci M. Chondrosarcoma of bone: the experience at the Istituto Ortopedico Rizzoli. *J Bone Joint Surg Am.* 1981;63:1248–1257.

14. Hakim DN, Pelly T, Kulendran M, Caris JA. Benign tumours of the bone: a review. *J Bone Oncol.* 2015;4:37–41.

15. Henderson ED, Dahlin DC. Chondrosarcoma of bone: a study of two hundred and eighty-eight cases. *J Bone Joint Surg Am.* 1963;45:1450–1458.

16. Jesus-Garcia R, Osawa A, Filippi RZ, Viola DC, Korukian M, de Carvalho Campos Neto G, Wagner J. Is PET-CT an accurate method for the differential diagnosis between chondroma and chondrosarcoma? *Springerplus.* 2016;5:236.

17. Jones GL, Bishop JY, Lewis B, Pedroza AD; MOON Shoulder Group. Intraobserver and interobserver agreement in the classification and treatment of midshaft clavicle fractures. *Am J Sports Med.* 2014;42:1176–1181.

18. Lee FY, Mankin HJ, Fondren G, Gebhardt MC, Springfield DS, Rosenberg AE, Jennings LC. Chondrosarcoma of bone: an assessment of outcome. *J Bone Joint Surg Am.* 1999;81:326–338.

19. Lichtenstein L, Jaffe HL. Chondrosarcoma of bone. *Am J Pathol.* 1943;19:553–589.

20. Marco RA, Gitelis S, Brebach GT, Healey JH. Cartilage tumors: evaluation and treatment. *J AmAcad Orthop Surg.* 2000;8:292–304.

21. Murphey MD, Flemming DJ, Boyea SR, Bojescul JA, Sweet DE, Temple HT. Enchondroma versus chondrosarcoma in the appendicular skeleton: differentiating features. *Radiographics.* 1998;18:1213–1237; quiz 1244–1215.

22. Nota SP, Braun Y, Schwab JH, van Dijk CN, Bramer JA. The identification of prognostic factors and survival statistics of conventional central chondrosarcoma. *Sarcoma.* 2015;2015:623746.

23. Ozaki T, Hillmann A, Linder N, Blasius S, Winkelmann W. Metastasis of chondrosarcoma. *J Cancer Res Clin Oncol.* 1996;122:629–632.

24. Ozaki T, Lindner N, Hillmann A, Rodl R, Blasius S, Winkelmann W. Influence of intralesional surgery on treatment outcome of chondrosarcoma. *Cancer.* 1996;77:1292–1297.

25. Rotondi MA, Donner A. A confidence interval approach to sample size estimation for interobserver agreement studies with multiple raters and outcomes. *J Clin Epidemiol.* 2012;65:778–784.

26. Schreuder HW, Pruszczynski M, Veth RP, Lemmens JA. Treatment of benign and low-grade malignant intramedullary chondroid tumours with curettage and cryosurgery. *Eur J Surg Oncol.* 1998;24:120–126.

27. Skeletal Lesions Interobserver Correlation among Expert Diagnosticians (SLICED) Study Group. Reliability of histopathologic and radiologic grading of cartilaginous neoplasms in long bones. *J Bone Joint Surg Am.* 2007;89:2113–2123.

28. Urrutia J, Zamora T, Yurac R, Campos M, Palma J, Mobarec S, Prada C. An independent inter- and intra-observer agreement evaluation of the AOSpine subaxial cervical spine injury classification system. *Spine (Phila Pa 1976).* 2015 Nov 30. (Epub ahead of print).

29. Urrutia J, Zamora T, Yurac R, Campos M, Palma J, Mobarec S, Prada C. An independent interobserver reliability and intraobserver reproducibility evaluation of the new AOSpine Thoracolumbar Spine Injury Classification System. *Spine (Phila Pa 1976).* 2015;40:E54–58.

30. Verdegaal SH, Brouwers HF, van Zwet EW, Hogendoorn PC, Taminiau AH. Low-grade chondrosarcoma of long bones treated with intralesional curettage followed by application of phenol, ethanol, and bone-grafting. *J Bone Joint Surg Am.* 2012;94:1201–1207.

31. Wang XL, De Beuckeleer LH, De Schepper AM, Van Marck E. Low-grade chondrosarcoma vs enchondroma: challenges in diagnosis and management. *Eur Radiol.* 2001;11:1054–1057.