Interobserver and Intraobserver Reliability of the Enneking Classification in Plain Radiographic Staging of Benign Bone Tumors of the Extremities in Patients Seen at the Philippine General Hospital

Bernardino B. Alpuerto II, MD and Edward H.M. Wang, MD

Department of Orthopedics, Philippine General Hospital, University of the Philippines Manila

ABSTRACT

Objective. To determine the interobserver and intraobserver reliabilities of the Enneking Classification system in staging benign bone tumors.

Methods. Photographs of traditional plain radiographs of 65 histologically benign tumors from the PGH Department of Orthopedics Tumor Registry were used in the study. Nine Orthopedic surgeons (three consultants, one fellow, and five senior residents) staged the tumors using the Enneking Classification based on radiographic tumor-host margins. The photographs were sent to the surgeons twice (batch 1 and 2), three months apart, for staging. The Fleiss and Cohen kappa statistics were used to determine interobserver and intraobserver reliabilities, respectively. This is a pilot study.

Results. There was only fair interobserver reliability of the Enneking Classification staging with Fleiss kappa of 0.38 and 0.26 for batches 1 and 2, respectively. Also, there was only moderate intraobserver reliability (Cohen kappa 0.48) for the staging. Moreover, there was also a relatively low intraobserver percent agreement (67%) among raters. In both reliabilities, the consultants/fellow group consistently showed better interobserver and intraobserver reliabilities compared to the residents.

Conclusion. The Enneking Classification in staging benign bone tumors had relatively low interobserver and intraobserver reliabilities. There was also a tendency of experienced orthopedic tumor consultants and senior residents to stage the same radiograph differently upon repeat testing.

Key Words: Enneking, tumors, reliability, interobserver, intraobserver

INTRODUCTION

Benign bone tumors of the extremities are relatively common entities. At the Philippine General Hospital, they make up 40% of all primary bone tumors.[1] These bone lesions include a variety of lesions with different biologic behavior. The more common types include giant cell tumor (GCT) and aneurysmal bone cyst (ABC) with incidences of 20% and 9%, respectively. In general, the prognosis is good for benign lesions but results of treatment vary according to tumor biology and the degree of aggressiveness.

Enneking proposed a staging system for benign and malignant bone tumors, wherein Arabic numerals (stages 1, 2, and 3) are designated for benign lesions and Roman numerals (stages I, II, and III) are used for the malignant lesions. The classification is based on the radiographic

characteristics of the tumor-host margin.[2] The Enneking staging labels the benign bone tumors into latent, active, or aggressive, depending on the radiologic findings, histology, and clinical findings. Latent or inactive lesions have clear, geographic tumor-host bone borders on radiographs. Patients are usually asymptomatic and these lesions can be incidental findings on routine radiographs. Active lesions have increased sizes but are still usually contained within the bone with some amount of endosteal expansion seen in radiographs. Patients can complain of some pain and tenderness over the mass. Aggressive lesions are larger and are often poorly contained within bone on radiographs. These patients present with more pain aside from a prominent soft-tissue mass. For malignant tumors, the Enneking surgical staging system takes into account the surgical grade (G, G1, G2), local extent (T, T1, T2), and presence/absence of metastasis (M0, M1).[2] This study only focuses on the Enneking staging system for benign bone tumors.

An ideal staging system should be practical, reproducible, and prognostic significance.[2,3] In a study done by Chan et al.,[4] the reliability of the Enneking and WBB classifications for staging primary spine tumors was assessed. Results indicated moderate interobserver reliability of both staging classification and near-perfect intraobserver reliability in terms of staging and guidance for treatment. Both benign and malignant spine tumors were included in that study.

Similar studies are also found in the literature. A study done by Wood et al. in 2005, where they assessed two thoracolumbar fracture classification systems used by multiple surgeons; emphasized that repeatability and reproducibility of classification systems are necessary for accurate communication between treating physicians.[5] In their study, the Denis and AO classification systems for thoracolumbar fractures were assessed by 19 fellowship-trained spine surgeons using radiographs and CT scan views of the fractures. Interobserver and intraobserver reliabilities were obtained using the Cohen kappa test and both showed only moderate reliability and repeatability.

Another study, but this time assessing only the reliability of the Cervical Spine Injury Severity Score was done by Anderson et al. in 2007.[6] This scoring system is based on consideration of both ligamentous and osseous injury determining the stability of the cervical spine wherein ≥7 score needing surgery. Their results showed excellent intraobserver and interobserver reliability and this scoring system can help in development of treatment algorithms for clinical trials.

Reliability of measurements is the extent to which clinicians agree in their ratings and not merely the extent to which their ratings are associated or correlated.[7] The calculation of reliability is based on the difference between how much agreement is actually present (observed agreement) compared to how much agreement would be expected to be present by chance alone (expected agreement). Kappa (κ) is a measure of this difference, standardized to lie on a -1 to 1 scale, where 1 is perfect agreement, 0 is exactly what would be

expected by chance, and negative values indicate agreement less than chance; potential systematic disagreement between observers.[8] To interpret the calculated value of κ, a scale can be used from 0 to 1 wherein <0 is labeled "less than chance agreement" and 1 is "almost perfect agreement."

Formulation of a treatment plan for benign bone tumors is in large part dependent on the Enneking staging system. Both surgical oncologic margins and the use of adjuvants are based on the stage assigned to the tumor (Table 1). The system, therefore, requires adequate agreement among different treating surgeons (interobserver reliability) to facilitate communication. It also requires intraobserver reliability or same surgeon consistency. Once the reliability and validity of a staging system have been established, the system can more easily provide for a standardized approach to treat these tumors. The Enneking classification system for benign bone tumors has been in use for over 30 years.[9] It has been validated but this was done with malignant tumors specifically in musculoskeletal sarcoma. As such, the study aims to determine the reliability of the classification but this time with benign bone tumors of the extremities in the local setting.

Table 1. Enneking stages and recommended surgical margins

Enneking Stage	Margin of Control
1	No management unless for decompression or stabilization
2	Intralesional excision ± local adjuvants
3	Marginal *en bloc* excision

MATERIALS AND METHODS

Procedure

Inclusion criteria

Pre-operative radiographs (AP and lateral views) of histologically determined benign bone tumors of the extremities were gathered from the PGH Tumor database.

Exclusion criteria

Radiographs of malignant bone tumors were not included, as well as post-operative radiographs of benign bone tumors.

Study duration

This study was done for around six to eight months. A total of sixty-five (65) traditional plain radiographs of the tumors (AP and lateral views) were photographed using a Canon Powershot A3300 IS 16-megapixel camera and compiled in flash drives together with handouts and answer sheets in 2 sets. The handouts contained instructions and a case example to clarify the methodology. The radiographs were arranged randomly. The 2nd set contained the same photographed x-rays and handouts but arranged in a different order.

The materials (65 photographed radiographs with instructions and answer sheets) were sent to nine orthopedic surgeons (3 PGH orthopedic tumor surgeons, 1 tumor fellow, and 5 PGH senior orthopedic residents). Each observer was tasked to stage all the 65 radiographs of the tumors using the Enneking classification system by marking the boxes in the answer sheets.

Once completed, the data were submitted to an independent authorized person (PGH Orthopedics tumor secretary) for filing. The same materials were re-sent to the same observers after 12 weeks in a different order to limit recall bias.

Nine orthopedic surgeons were able to return the answer sheets for the 1st set however only eight surgeons complied for the 2nd set.

Sample Size Calculation

A statistician from the Institute of Clinical Epidemiology, National Institutes of Health, UP Manila was consulted regarding sample size calculation.

Sample Size Determination

For a simple random sample to be taken from a population in which the proportion of patients with benign tumor is the statistic to be estimated and the population size of patients with bone tumor is unknown, the formula used to compute for the sample size is shown below:

$$n = (PQ) (Z_{a/2}/B)^2$$

where:
P = proportion of patients with benign tumor; (0.3971)
Q = 1-P; (0.6029)
B = maximum allowable error; (0.1)
α = level of significance (0-.05); and
$Z_{\alpha/2}$ = standard normal variate (1.645)

The proportion of patients with benign tumors was obtained from the local publication of Wang et al. in 2007. The sample size calculated was 65. Radiographs of 65 patients with benign bone tumors in the extremities were obtained from the PGH Tumor database. The selected radiographs for the study were all good quality with no film distortions and no film blemishes. All of the patients already have a histologic diagnosis of a benign bone tumor.

Statistical Analysis

The Fleiss κ statistic was used for measuring interobserver reliability of the Enneking Classification system; while for intraobserver reliability, the Cohen κ was used. To main-tain consistent nomenclature when describing the relative strength of agreement associated with kappa statistics, the following labels were assigned to the corresponding ranges of kappa as shown in Table 2.[10]

Table 2. Percentage of Agreement at a variety of κ Statistics Levels (Landis and Koch)

κ Value	Strength of Agreement
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.80-1.00	Almost Perfect

RESULTS AND DISCUSSION

Histology of Radiographs

Sixty-five radiographs of benign bone tumors of the extremities were used in this reliability study. The histology of the bone tumors is shown in Table 3. The aneurysmal bone cyst has the highest number of radiographs followed by giant cell tumors.

Table 3. Histology of radiographs of benign bone tumors of extremities

Histology	Number of radiographs
Aneurysmal Bone Cyst	27
Giant Cell Tumor	18
Ostechondroma	4
Fibrous dysplasia	6
Chondroblastoma	4
Enchondroma	2
Unicameral bone cyst	4
Total	**65**

Sample radiographs with Enneking classification staging from stage 1 to 3 are shown in Figures 1 to 3.

Interobserver Reliability of Enneking Classification staging for Benign Bone Tumors

The interobserver reliability of Enneking classification staging using radiographic characteristics of tumor-host margins has a Fleiss κ statistics of 0.38, for the 1st batch and 0.26 for the 2nd batch. Both have a fair agreement based on the Landis and Koch κ statistics levels (Table 4).

The Fleiss κ for consultants and fellow group has a value of 0.48 (moderate agreement) compared to 0.36 (fair agreement) for the residents' group. The same trend was observed (0.54 – moderate agreement vs 0.17 – slight agreement) for the 2nd batch of radiographs.

Intraobserver Reliability of Enneking Classification staging for Benign Bone tumors

The intraobserver reliability of Enneking staging has a Cohen κ statistics of 0.48 (moderate agreement) with only 67% agreement. The consultants and fellow group have a Cohen κ of 0.64 (substantial agreement) compared to the residents' group (0.39 – fair agreement). The percent agreement (80%) for the consultant/fellow group was also higher compared to the residents' group (58%).

DISCUSSION

The Enneking classification staging for benign bone tumors has been in use for over 30 years but has never been validated. A classification system must have both interobserver and intraobserver reliability studies to provide a more standardized approach in treatment and also for prognostication. This study is the first to attempt an evaluation of the interobserver and intraobserver reliabilities of the Enneking classification staging for benign bone tumors. To our knowledge, there are no other similar studies published in the English literature.

The staging system for benign bone tumors considers the tumor grade, site, and presence/absence of metastasis; but for this study, only the radiographic characteristics of the tumor-host margins were considered.

Figure 1. *Enneking Classification Stage 1:* Well-demarcated borders; clear geographic borders.

Table 4. Results of the Interobserver reliability study for Enneking classification staging of benign bone tumors

	Set 1 (κ)	Set 2 (κ)
Stage		
1	0.25	0.16
2	0.24	0.17
3	0.67	0.42
Combined	0.38	0.26
Consultants/Fellow	0.48	0.54
Residents	0.36	0.17

Table 5. Results of the Intraobserver reliability study for Enneking classification staging of benign bone tumors

Rater	κ	Percent Agreement
Combined	0.48	67%
Consultants/Fellow κ	0.64	80%
Residents κ	0.39	58%

Figure 2. *Enneking Classification Stage 2:* Indistinct borders; increased sizes; usually contained within the bone with endosteal expansion.

Figure 3. *Enneking Classification Stage 3:* Indistinct borders, increased sizes, and poorly contained within the bone.

The summary of Fleiss κ statistics is shown in Table 4. There was only fair agreement (κ = 0.38) in interobserver reliability of the Enneking classification. Kappa statistics were also computed for each Enneking stage. For tumors labeled as stage 1 and stage 2, the kappa statistics were 0.25 and 0.24 (fair agreement), respectively. For stage 3 labeled tumors, the kappa statistic was 0.67 (substantial agreement). These results imply that the orthopedic surgeons were more able to agree with the staging if the radiographic characteristics of the benign tumors fall under the Enneking stage 3. These tumors were described in radiographs to have an indistinct ragged permeative interface with the adjacent bone, incomplete attempts at containment by reactive bone, cortical destruction, endosteal buttresses, and periosteal Codman's triangles, and rapid soft tissue extension.[11] There will likely less disagreement with these radiographic characteristics, allowing a more consistent staging of Enneking 3. Only eight orthopedic surgeons were able to submit the answer sheets for the second set of radiographs. One orthopedic consultant was not able to submit back the answer sheet upon the designated time for data analysis. Similar photographs of the radiographs were sent after three months but in a different order. The computed Cohen κ statistics for intraobserver reliability was only 0.48 (moderate agreement). This meant that there was a significant change in the staging of the surgeon of the radiographs after three months. As shown in Table 5, there is a trend of more changes in staging between Enneking stage 1 and Enneking stage 2; implying the effort in committing on what the stage of the bone tumor using the radiographic characteristics of the Enneking classification staging.

The kappa statistics were computed for consultants and fellow separately from the residents; and showed that for both the interobserver and intraobserver reliabilities; the agreement was higher in the consultants/fellow group than in the residents (Tables 4 and 5). This indicates that although the Enneking classification describes each radiographic stage adequately, the experience of the orthopedic surgeon is also vital. The three orthopedic tumor fellowship-trained consultants have at least 10 years of experience treating tumor cases; the tumor fellow was on his last three months of a 1-year tumor orthopedic subspecialty fellowship training; the five senior residents were in their 4th year of orthopedic residency training. With enough experience and abundant radiographs previously reviewed, a tumor surgeon can extend beyond the radiographic characteristics to stage the tumor. Based on the radiographs also showing the location of the tumor, the grade and possibly the histology of the lesion can be inferred which can help in staging since there are already representative bone tumor histology types for each Enneking stage. For example, an osteoid osteoma usually falls under stage 1 while an aneurysmal bone cyst is often stage 2; and giant cell tumor is either stage 2 or 3. Though they are already senior residents; they lack this experience that might have resulted in the decreased interobserver and intraobserver

reliability of Enneking classification. On the other hand, though the Enneking staging is simple and easy to teach, it is subject to non-objective or non-measurable parameters merely based on how the surgeons assess the characteristics of the tumor-host margins on radiographs. Moreover, those parameters can be affected by the quality of the picture as well as the clarity/exposure of the radiographs.

No similar literature regarding interobserver and intraobserver reliabilities of the Enneking classification staging for benign bone tumors was found. Moreover, while this study shows a reliability level of the Enneking classification system that is lower than expected or hoped for; there are at present no other similar classification systems describing benign bone lesions in stages were found.[12] Drumond et al reported that the Enneking staging system was very effective in determining prognosis and for planning the surgery for benign bone tumors and tumor-like lesions with a 95.2% agreement. However, in his retrospective study, he used a combination of plain radiographs, computed tomography, nuclear magnetic resonance scintigraphy, and arteriography to determine the Enneking stage of the benign bone tumors. The Lodwick classification is similar to the Enneking staging, but it mainly classifies bone lesions into types depending on the radiographic characteristics whether it appears benign or malignant.[13]

The Enneking staging system was developed in a time with only simple radiographs were available and remained unmodified and unvalidated statistically until now. Given the results of our study, it is advisable and an opportune time to reevaluate and possibly modify the Enneking staging system incorporating newly available imaging modalities.

CONCLUSION

There were relatively low interobserver and intraobserver reliabilities of the Enneking classification staging for benign bone tumors based on radiographic characteristics. This raises questions about whether it can be used alone to determine preoperatively the appropriate treatment for the patient and also the recommended surgical margins. Based on our results, it is recommended that the Enneking system not be used alone to classify benign bone tumors. The orthopedic surgeon should interpret the aggressiveness of the tumor based on a combination of this system, good physical examination, and other imaging studies that may be available on consultation.

Limitations

The lack of a standardized x-ray machine resulting in differential exposures of the radiographs that caused varying qualities of the pictures. The radiographs were only obtained from the PGH Tumor database which was cases compiled through the years. Though it cannot be assessed objectively, the authors personally selected the best radiographic films of the tumors in terms of factor exposure and quality that

would allow adequate staging and even appropriate surgical decision making. To reduce observer bias and enhance generalizability, orthopedic tumor consultants, tumor fellow, and senior residents were selected as raters for the study. Though explanations and instructions were sent to the raters, it was assumed that they already know the Enneking Classification staging of benign bone tumors. The range and extent of the experience of the raters were no longer formally verified and evaluated.

The Enneking classification of benign bone tumors also includes or takes into consideration the grade of the tumor, location, and presence/absence of metastasis; but for this study, the focus was to use the radiographic tumor-host margins to stage the benign bone tumors. Most of the time, referrals to orthopedic surgeons mainly involve radiographs and it is not uncommon that the surgeons were not able to assess the patient clinically or histologically preoperatively; making the radiographic evaluation of the stage of the tumor essential for the treatment plan.

Statement of Authorship

Both authors participated in data collection and analysis, and approved the final version submitted.

Author Disclosure

Both authors declared no conflicts of interest.

Funding Source

No funding support.

REFERENCES

1. Wang EHM and de Dios AV. Bone Tumors in Filipinos. A collection of 876 cases from the University of the Philippines, Philippine General Hospital. The Bookmark Incorporated, Philippines. 2007.
2. Jawad SP and Scully SP. Classifications in Brief: Enneking Classification: Benign and Malignant Tumors of the Musculoskeletal System. Clin Orthop Relat Res. 2010 Jul;468(7):2000-2.
3. Patel AA, Vaccaro AR, Albert TJ, Hilibrand AS, Harrop JS, Anderson GD, et al. The Adoption of a New Classification System: Time dependent variation in Interobserver Reliability of the Thoracolumbar Injury Severity Score Classification System. Spine (Phila Pa 1976). 2007 Feb 1;32(3):E105-10.
4. Chan P, Boriani S, Fourney DR,Biagini R, Dekutoski MB, Fehlings MG, et al. An assessment of the reliability of the Enneking and Weinstein-Boriani-Biagini Classifications for staging of Primary Spinal tumors by the Spine Oncology Study Group. Spine (Phila Pa 1976). 2009 Feb 15;34(4):384-91.
5. Wood KB, Khanna G, Vaccaro AR, Arnold PM, Harris MB, Mehbod AA. Assessment of two thoracolumbar fracture classification systems as used by Multiple surgeons. J Bone Joint Surg Am. 2005 Jul;87(7):1423-9.
6. Anderson PA, Moore TA, Davis KW, Molinari RW, Resnick DK, Vaccaro AR, et al. Cervical Spine Injury Severity Score: Assessment of Reliability. J Bone Joint Surg Am. 2007 May;89(5):1057-65.
7. Sim J and Wright CC. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. Phys Ther. 2005 Mar;85(3):257-68.
8. Vierra AJ and Garrett JM. Understanding Interobserver Agreement: The Kappa Statistic. Fam Med. 2005 May;37(5):360-3.
9. Enneking WF, Spanier SS, Goodman MA. Current Concepts Review: The Surgical Staging of Musculoskeletal Sarcoma. J Bone Joint Surg Am. 1980 Sep;62(6):1027-30.
10. Landis RJ and Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977 Mar;33(1):159-74.
11. Enneking WF. A System of Staging Musculoskeletal Neoplasms. Clin Orthop Relat Res. 1986 Mar;(204):9-24.
12. Drumond, Jose Marcos. Efficacy of the Enneking staging system in relation to treating benign bone tumors and tumor-like bone lesions. Rev Bras Ortop. 2010;45(1):46-52
13. Clinical MRI. Pictorial Essay: Benign and Malignant Bone Tumors Radiologic Diagnosis and Imaging Features. [Internet] 2013 [cited November 2014]. Available from http://clinical-mri.com/pictorial-essay-benign-and-malignant-bone-tumors-radiological-diagnosis-and-imaging-features/.

APPENDICES

Department of Orthopedics Tumor Section

Philippine General Hospital

Dear Doctor:

I am currently conducting a research entitled, "Interobserver and intraobserver reliability of the Enneking Classification in staging benign bone tumors of the extremities in PGH patients". This study aims to validate the current classification system used in staging benign bone tumors in PGH; whether there is adequate agreement and the results are reproducible among orthopedic surgeons.

In line with this, I am requesting your participation by using the Enneking classification in staging a set of radiographs with benign bone tumors. I have provided simple instructions and answer sheets for this purpose.

Another set of radiographs will be sent after 3 months for you to stage; this time to determine intraobserver reliability of Enneking Classification.

Thank you in advance in participating in this study. There will be no compensation monetary or in kind in your participation in this study.

Bernardino B. Alpuerto II, MD
Orthopedic Resident
Philippine General Hospital

_____

Conforme

I have understood the purpose of this study and I willingly agreed to participate.

Name: _____

Signature: _____

Date: _____

Instructions

Each envelope contains the following:
- A flashdrive with pictures of radiographs
- An answer sheet
- An instruction sheet
- A consent form
- An example case

1. Insert the flashdrive in the USB port of your computer/laptop.
2. Within the drive are numbered pictures of radiographs of benign bone tumors. Do not rearrange the pictures.
3. View the numbered radiographs and decide the Enneking tumor stage.
4. Check the appropriate numbered box of the tumor stage.
5. Once completed, return the flashdrive and other sheets in the envelope and submit it to the Tumor Section secretary.

Table 1. Enneking staging for benign musculoskeletal tumors based on radiographic characteristics of tumor host margins

Stage	Description
Latent	Well-demarcated borders; clear geographic borders
Active	Indistinct borders; increased sizes; usually contained within the bone with endosteal expansion
Aggressive	Indistinct borders; increased sizes and poorly contained within the bone

Umar et al. 2010. In Brief: Enneking Classification: Benign and Malignant Tumors of Musculoskeletal System. Clin Orthop Relat Res 468: 2000-2002

Sample Case

Radiograph number	Enneking Stage		
	1	2	3
1			✓
2			

ANSWER SHEET

Set number: _____

Rater's code number: _____ [   ] consultant [   ] resident

Radiograph number	Enneking Stage		
	1	2	3
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			

Radiograph number	Enneking Stage		
	1	2	3
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			

Radiograph number	Enneking Stage		
	1	2	3
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			
65			