

Inter-rater reliability of the radiographic assessment of simple bone cysts

S. Cho¹
R. Yankanah¹
P. Babyn²
J. Stimec³
A. S. Doria³
D. Stephens⁴
J. G. Wright¹

Abstract

Purpose To develop and evaluate the reliability of an explicit set of parameters and criteria for simple bone cysts (SBCs) and evaluate the reliability of single *versus* serial chronological reading methods.

Methods Radiographic criteria were developed based on the literature and expert consensus. A single anteroposterior/lateral radiograph from 32 subjects with SBC were evaluated by three radiologists. A second reading was then conducted using revised criteria including a visual schematic. In the third reading the same images were assessed but radiologists had access to images from two additional time points. Inter-rater reliability was assessed after each reading using kappa (κ) and percentage agreement for categorical and binary parameters and intra-class correlation coefficient (ICC) for continuous parameters.

Results Parameters that were revised with more explicit definitions including the visual schematic demonstrated consistent or improved inter-rater reliability with the exception of continuous cortical rim present and cyst location in the metaphysis and mid-diaphysis. Cortical rim displayed only slight reliability throughout ($\kappa = -0.008$ to 0.16). All other categorical parameters had a percentage agreement above 0.8 or a moderate ($\kappa = 0.41$ to 0.60), substantial ($\kappa = 0.61$ to 0.80) or almost perfect inter-rater reliability ($\kappa = 0.81$ to 1.0)

in at least one reading. All continuous parameters demonstrated excellent inter-rater reliability (ICC > 0.75) in at least one reading with the exception of scalloping (ICC = 0.37 to 0.70). Inter-rater reliability values did not indicate an obviously superior method of assessment between single and serial chronological readings.

Conclusion Explicit criteria for SBC parameters used in their assessment demonstrated improved and substantial inter-rater reliability. Inter-rater reliability did not differ between single and serial chronological readings.

Level of Evidence: Not Applicable

Cite this article: Cho S, Yankanah R, Babyn P, Stimec J, Doria AS, Stephens D, Wright JG. Inter-rater reliability of the radiographic assessment of simple bone cysts. *J Child Orthop* 2019;13:226-235. DOI: 10.1302/1863-2548.13.180140

Keywords: simple bone cyst; unicameral bone cyst; bone cyst; inter-rater reliability; radiography

Introduction

Simple bone cysts (SBC) otherwise known as unicameral bone cysts are the most common benign bone lesions in children and adolescents.^{1,2} Patients with SBCs usually present with fracture or pain in the affected limb.³⁻⁵ SBCs have a high probability of recurrence and repeat fracture after treatment,^{3,6-10} and may even cause discrepant limb lengths or angular deformity.^{1,4,7,11} While benign, SBCs can have a life-long impact on children through risk of fracture and activity restriction, thus, clinicians need reliable measures for their assessment.^{8,10,12}

The radiographic features of SBC can include those used for diagnosis such as the location of the cyst, prognosis such as the cyst index,¹³ or as an outcome of treatment, for instance, the Neer classification.¹⁴ Other SBC features are also consistent with various benign bone lesions, such as cortical reaction. However, few of these parameters have been shown to be reliable and the absence of specific criteria may partially explain the conflicting findings as to which characteristics can be used to predict healing or fracture.^{12,13}

Beyond the criteria used to evaluate an evolving disease such as SBC, whether radiographs should either be interpreted individually without review of prior images or assessed in a chronological series has received little

¹ Department of Surgery, The Hospital for Sick Children, Toronto, Ontario, Canada

² Department of Medical Imaging, University of Saskatchewan and Saskatoon Health Region, Saskatoon, Saskatchewan, Canada

³ Department of Diagnostic Imaging, The Hospital for Sick Children, Toronto, Ontario, Canada

⁴ Research Institute, The Hospital for Sick Children, Toronto, Ontario, Canada

attention in the orthopaedic literature. While serial chronological readings may introduce bias from prior examinations,¹⁵ it may be advantageous to have an illustration of the chronological progression of the disease which could lead to assessments more sensitive to change.^{15,16}

The purpose of this study was: a) to evaluate the reliability of radiographic criteria used in diagnosis, prognosis or outcome of SBC as a typical benign bone lesion; and b) to evaluate the reliability of a single *versus* serial chronological evaluation of radiographs for SBCs.

Materials and methods

Patient sample

The patient sample consisted of 32 patients aged four to 14 years (median ten years) with a radiographic diagnosis of a SBC. All participants were previously enrolled in a multi-centre trial for the treatment of a SBC.¹⁷ For each study subject, an anteroposterior (AP) and lateral (Lat) radiograph of the cyst at baseline, one year and two years post-treatment were obtained and de-identified. Approval for this study was granted by the institutional research ethics board of the Hospital for Sick Children, Toronto, Ontario, Canada.

Development of radiographic parameters

A literature search using MEDLINE was conducted to collate radiographic characteristics of SBCs. The following search string was used: bone cysts/and (unicameral or multicameral or simple), radiography, treatment outcome, treatment failure and aneurysmal bone cysts. The search was limited to the English language and filtered to include humans and children 0 to 18 years old. Redundant parameters describing similar characteristics were consolidated or eliminated. Parameters for characterization of bone cysts were defined either by consensus of authors or by modified criteria from existing definitions in the literature (Table 1). In some cases, variables were assigned reference images (Fig. 1) and those parameters related to continuous data such as cortical thinning were divided into categorical scales such as 0% to 25%, 25% to 50%, etc. Factors were grouped into those used for diagnosis, prognosis and outcome assessment of SBC.

Radiographic assessment

The readers involved were three board-certified paediatric musculoskeletal radiologists with 29, seven and 13 years of experience after training, respectively (PB, ASD, JS). Each radiograph was assigned a subject ID and images were read using picture archiving and communications system workstations with the default

calibration settings and measurement tools available in clinical practice. The radiologists had no prior discussion regarding measurement or specific instructions other than the criteria provided. Data was recorded on a paper case report form and later inputted into a secure RED-Cap database.¹⁸

First reading

AP/Lat images from one of the three time points available for each of the 32 patients was randomly selected and reviewed by the three radiologists. After an initial analysis, with the assistance of one of the radiologists (ASD) more precise definitions were added to the criteria with particular attention to those with poor reliability (Table 1). A visual schematic illustrating how to classify various cystic features was also created (Fig. 2).

Second reading

Upwards of one year after the first reading, the two radiologists (PB, JS) not involved in the revisions to the parameters evaluated the same 32 AP/Lat radiographs using the revised criteria, case report form and the visual schematic.

Third reading

At least two weeks after the second reading, the two radiologists (PB, JS) evaluated the same 32 radiographs using the criteria. However, each radiologist had the ability to view all three images for the corresponding patient simultaneously and had knowledge of their chronological order (serial chronological method).

Statistical analysis

Based on the number of radiologists per reading, sample size calculations yielded estimates of 32 patients required for the first reading and 19 required for the second and third reading.¹⁹ All inter-reader reliability was determined using the intra-class correlation coefficient (ICC) for continuous parameters and the kappa statistic (κ) and percentage agreement for categorical or binary parameters. ICC values range from 0 to 1 and values above 0.75 were considered to have excellent reliability.²⁰ Kappa values range from -1 to 1²¹ and were categorized as poor ($\kappa < 0$), slight ($\kappa = 0$ to 0.20), fair ($\kappa = 0.21$ to 0.40), moderate ($\kappa = 0.41$ to 0.60), substantial ($\kappa = 0.61$ to 0.80) and almost perfect ($\kappa = 0.81$ to 1).²² The kappa statistic tends to penalize the reliability of data with an imbalanced distribution that is heavily skewed towards positive agreement,²³⁻²⁷ as agreement attributed to chance is then assumed to be high which then lowers the kappa value.²⁸ Therefore, percentage agreement was also used as a primary measure of reliability in this study.

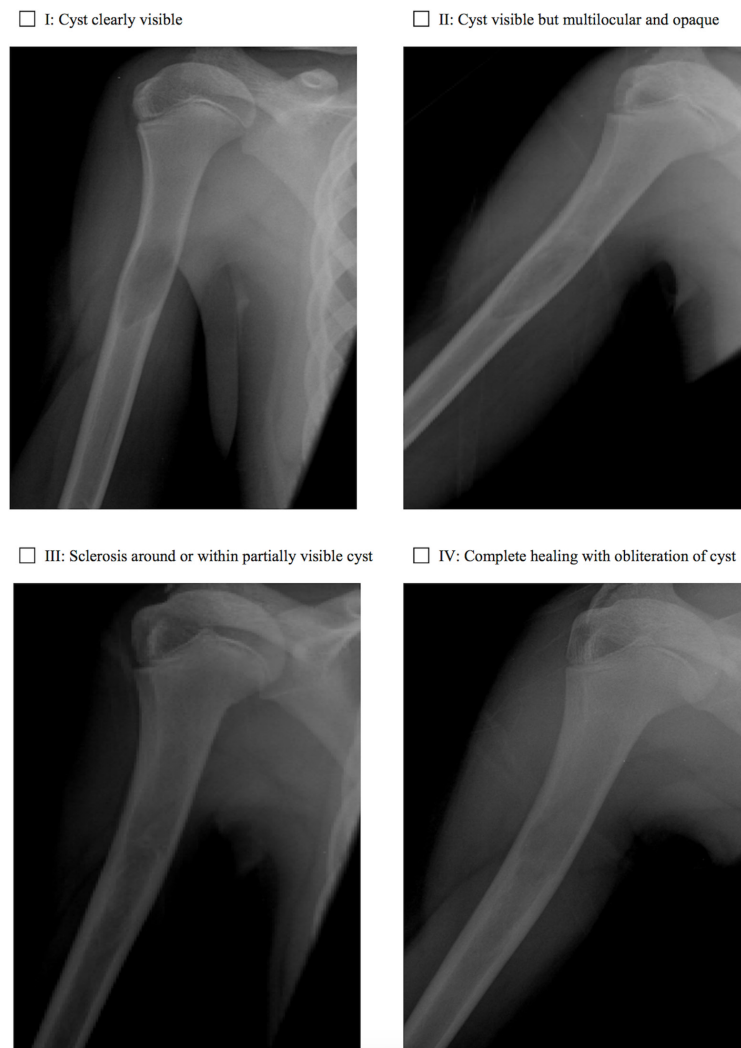


Fig. 1 Product scale developed to represent the cyst grade spectrum.

Results

Out of the 32 patients, 22 (69%) were male and ten (31%) were female. There were 24 (75%) and eight (25%) SBCs located in the humerus and femur, respectively. A summary of the resulting reliability values for bone measurements and each of the parameters can be found in Table 2. All categorical parameters demonstrated moderate to almost perfect kappa values during at least one of the readings with the exception of the percentage of continuous cortical rim and cyst location, which consistently had a poor or slight kappa value. The percentage of scalloping present around the cyst border was the only continuous parameter to consistently demonstrate less than excellent reliability.

Within the diagnostic parameters, cyst location and periosteal reaction demonstrated variability in kappa

values, however, both had percentage agreement values above 0.8 in at least one of the readings. Loculation had fair to substantial reliability across readings. Prognostic parameters based on continuous data demonstrating consistently excellent reliability included cyst dimensions, cyst volume and tubulation measurements. Cyst activity and indices demonstrated less than excellent reliability in the first reading only, otherwise they also showed excellent reliability. Prognostic parameters based on categorical data including the presence and absence of tubulation, whether the tubulation caused expansion to the bone and the presence and absence of scalloping demonstrated slight to substantial kappa values but had percentage agreement values above 0.8 across the second and third readings. The outcome parameter of fracture had fair to substantial kappa values but percentage agreement values above 0.8. Cyst grade had fair to moderate reliability.

Table 1 Diagnostic, prognostic and outcome parameters and their definitions, criteria and interpretation when present in simple bone cysts

Parameter	Definition and interpretation	Initial radiographic criteria	Revised radiographic criteria
Diagnostic parameters			
Cyst location ^{4*}	Location of cyst within the bone. Are usually located in the proximal humerus or femur. ⁴	Check all that apply: 1) metaphyseal (in flare of bone) 2) epiphyseal (across physis) 3) diaphyseal (below flare of bone) 4) proximal 5) middle 6) distal.	Check all that apply: 1) metaphysis 2) epiphysis 3) proximal diaphysis 4) middle diaphysis 5) distal diaphysis.
Loculation ^{4*}	Presence of well corticated cavities. Unilocular: having a single cavity. Multiloculated: having many small cavities, a greater indication of recurrence. ^{7,8,37}	Number of loculations (well corticated cavities) present: 1) unilocular 2) 2 to 4 3) 5 to 10 4) > 10.	Number of intra-cystic loculations with a continuous contour present: 1) unilocular 2) 2 to 4 3) 5 to 10 4) > 10.
Cortical rim [*]	Cortical bone surrounding the cyst. A thickened continuous border is indicative of healing. An osteolytic rim is indicative of a higher risk of recurrence. ^{7,29,38} Incorporates radiodensity, zone of transition and geographic or permeative borders. A continuous cortical rim implies a greater percentage of radiodense rim, a sharper zone of transition and a geographic border.	Continuous cortical rim around cyst: 1) absent 2) 1% to 24% 3) 25% to 49% 4) 50% to 74% 5) 75% to 99% 6) 100%.	Continuous cortical rim around the cyst (% of total circumference): 1) absent 2) 1% to 24% 3) 25% to 49% 4) 50% to 74% 5) 75% to 99% 6) 100%.
Periosteal reaction ³⁹	Callus formation of thick, dense, wavy uniform new bone resulting from irritation or injury. ³⁹ Presence may be indicative of the existence of fracture. ³⁹	1) Absent 2) present.	
Bone measurements [*]	Dimensions of the bone in which the cyst is located.	1) Length of bone (middle of joint to middle of joint) (AP) 2) width of nearest physis (AP).	1) Length of bone (from level of proximal extent of joint to level of distal extent of joint) (AP) 2) thickness/width of nearest physis to cyst (AP) 3) diameter of diaphysis (AP).
Prognostic parameters			
Cyst activity ^{6,29*}	Distance of cyst to physis. Active cysts may increase fracture risk ^{1,37} . <i>Active:</i> within 1.0 cm to physis ²⁹ <i>Inactive:</i> > 1.0 cm to physis ²⁹	Smallest distance from middle of physis to top or bottom of cyst	Shortest distance from middle of the nearest physis to the outermost top or bottom margin of the cyst
Cyst volume [*]	Cyst length × width × depth Increase in volume is indicative of recurrence and potential fracture risk. ^{7,41}	Cyst measurements: 1) longest length (AP) 2) widest width (AP) 3) widest depth (AP).	Cyst measurements (from inner cortex to inner cortex): 1) longest length (from level of top to level of bottom of cyst) (AP) 2) widest width (AP) 3) widest depth (Lat). For chronological reading only; compared with most recent radiograph, how has the cyst volume changed?: 4) stable 5) decreased 6) increased 7) cyst absent.
Tubulation ^{40*}	Lateral growth and remodelling of the metaphysis into a tubular shape. ⁴⁰ Widening of the bone beyond normal diameter. ⁴ Wide metaphysis: indicative of non-healing. ⁴⁰	Defined as widening of the bone beyond normal diameter: 1) absent 2) present 3) unable to measure. Only if tubulation present: 4) widest width in area of tubulation 5) expected bone width in area of tabulation.	Defined as widening of the bone beyond normal diameter: 1) absent 2) present 3) unable to measure due to fracture. Only if tubulation present: 4) has the cyst expanded the diameter of the bone 5) widest width in area of tubulation 6) expected diameter of bone at level of greatest width.

Table 1 Cont.

Parameter	Definition and interpretation	Initial radiographic criteria	Revised radiographic criteria
Scalloping*	Curves along the edges of the cyst border due to thinning of the wall. ⁴ Is indicative of recurrence and non-healing. Incorporates endosteal thickening, or the thickness of the endosteal border.	Defined as thinning of cortex, measure based on two views: 1) absent 2) present 3) percentage of overall cyst border that is scalloped.	Defined as thinning of the cortex, measure based on two views: 1) absent 2) present 3) percentage of total circumference of the cyst.
Cyst index ¹³	Index is an assessment of the mechanical resistance of the cyst wall, with lower cyst indices associated with fracture and recurrence. ^{1,8,10,13,34} Based on the original equation, patients with a cyst index of > 5 mm and cortical thickness of < 1 mm are at greater risk of fracture. ⁴² As many cysts do not have a trapezoidal area which is required for the original equation, two modified cyst indices were developed and used.	1) Cyst length 2) cyst width 3) cyst depth 4) the narrowest diameter of the diaphysis on AP film (from outer cortex to outer cortex). Cyst index equations: $\text{Cyst index}^{13} = \frac{\text{cyst area}}{\text{diameter of diaphysis}^2}$ $\text{Cyst index}^{13} = \frac{\text{cyst area}}{\text{diameter of diaphysis}^2}$ $\text{Cyst index} = \frac{\text{cyst length} \cdot \text{cyst width} \cdot \text{cyst depth}}{\text{diameter of diaphysis}^2}$	
Outcome parameters Cyst grade ¹⁴	Cyst healing. Grade III and Grade IV cysts are considered to be satisfactorily healed. ³ Incorporates translucency and radiolucency which is equivalent to opacity.	Cole modification of the Neer Classification: ³ 1) Grade I – clearly visible 2) Grade II – visible but multilocular and opaque 3) Grade III – sclerosis around or within the partially visible cyst 4) Grade IV – complete healing with obliteration of the cyst. A product scale with an image representative of each grade was provided for visual reference during assessment (Fig. 1).	
Fracture	A break in the continuity of the bone as seen on the radiograph. Indicative of cyst recurrence.	1) Absent 2) present.	

*parameters that were illustrated on the visual schematic created for the second reading

AP, anteroposterior; Lat, lateral

Bone measurements consistently demonstrated excellent reliability with the exception of the width of the nearest physis in the first reading.

Overall, the second reading using the revised criteria and schematic displayed either consistent (difference < 0.10) or higher reliability values compared with the first reading for all continuous parameters and for all categorical parameters with the exception of the location in the metaphysis and mid-diaphysis and cortical rim for percentage agreement. Using the schematic and revised criteria, loculations, cyst activity, widest width of tubulation, expected bone width in tubulated area, percentage of scalloping around the cortical rim and bone measurements all demonstrated improved reliability between the first and second reading.

Comparing single and serial-chronological methods of reading

Between the single and serial chronological series methods, inter-rater reliability values did not indicate an obviously superior method. Parameters for which kappa

values and percentage agreement had improved by more than 0.2 (20%) were cyst location in the metaphysis and mid-diaphysis. Number of loculations had a kappa and percentage agreement value that decreased by more than 0.2 (20%). Continuous parameters had ICC values that remained stable across methods.

Discussion

Reliable radiographic assessment is essential for the diagnosis, prognosis and outcome of benign bone lesions. Although parameters such as the cyst grade, index and activity had explicitly defined criteria in the literature,^{3,13,14,29} the remainder of SBC parameters did not. This study initially evaluated the inter-rater reliability of parameters based on descriptions that were available, and after a schematic and precise criteria were added, reliability generally improved. Thus, the translation of these descriptive characteristics into specific binary categories, ordinal scales and quantitative measurements depicted in the

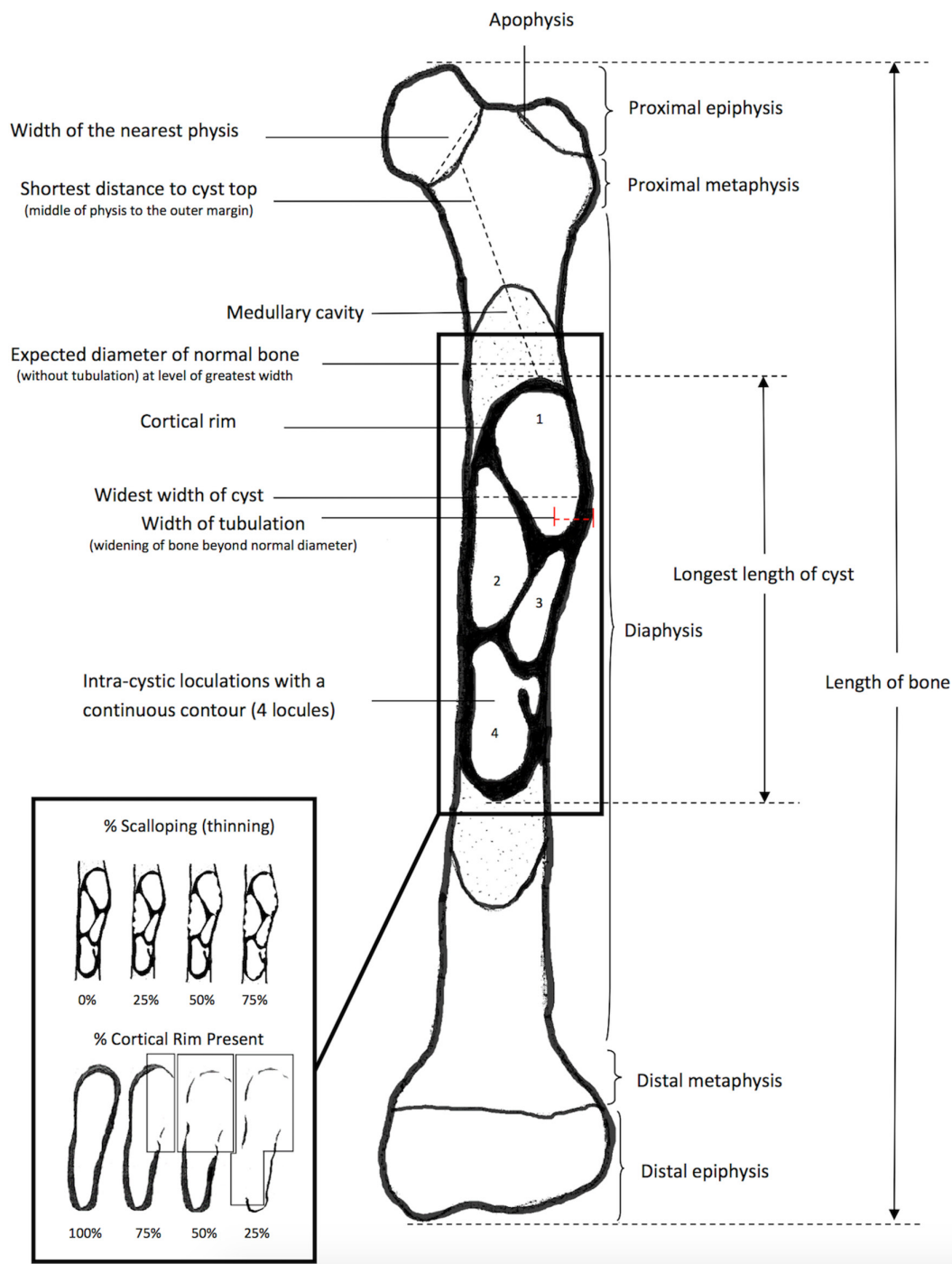


Fig. 2 Visual schematic developed to illustrate criteria for radiologists in the second and third readings.

revised criteria should guide clinicians when evaluating benign bone lesions such as SBCs.

Many of the SBC parameters that were evaluated in this study are also relevant to other bone lesions.³⁰⁻³⁶ In contrast to prior studies, one notable trend is that the kappa values for the categorical parameters in our study displayed

greater variance across readings than those reported in prior literature.^{30-33,35} The likely explanation is that whereas prior studies almost exclusively used binary scales (i.e. absent or present), this study used ordinal scales. Multiple categories provide clinicians with more specific information but may affect reliability.

Table 2 Kappa (κ), percentage agreement (%) and intra-class correlation coefficient (ICC) values for all parameters. Single hyphens indicate value not calculated due to data classification as continuous or categorical. Dashed lines are indicative of a parameter that was irrelevant to that particular stage of reading

Parameter	First reading				Second reading				Third reading							
	n	%	κ	ICC	95% CI	n	%	κ	ICC	95% CI	n	%	κ	ICC	95% CI	
Cyst location	Metaphysis	32	0.83	0.66	-	0.38 to 0.95	32	0.59	0.19	-	-0.14 to 0.53	32	0.91	0.81	-	0.61 to 1.00
	Epiphysis	32	0.97	0.01	-	-0.99 to 1.0	32	0.88	-0.067	-	-0.13 to -0.0015	32	0.97	0.00	-	0.0 to 0.0
	Prox. diaphysis	32	0.75	0.21	-	-0.11 to 0.53	32	0.84	-0.053	-	-0.14 to 0.033	32	0.94	0.63	-	0.16 to 1.00
	Mid diaphysis	32	0.85	0.70	-	0.41 to 0.98	32	0.63	0.26	-	-0.061 to 0.58	32	0.88	0.75	-	0.53 to 0.97
	Dist. diaphysis	32	0.97	0.01	-	-0.99 to 1.0	32	0.97	0.00	-	0.0 to 0.0	32	0.97	0.00	-	0.0 to 0.0
Loculation	Number of loculations	31	0.48	0.29	-	0.16 to 0.42	32	0.66	0.65	-	0.45 to 0.84	32	0.34	0.41	-	0.22 to 0.59
	% continuous cortical rim	32	0.30	0.16	-	0.06 to 0.26	32	0.094	0.005	-	-0.005 to 0.016	32	0.00	-0.008	-	-0.064 to 0.049
Periosteal reaction	Present/absent	32	0.92	0.66	-	0.45 to 0.87	32	0.84	0.45	-	0.047 to 0.86	32	0.66	0.068	-	-0.14 to 0.28
	Distance from physis to cyst top	29	-	-	0.73	0.56 to 0.85	31	-	-	0.89	0.78 to 0.95	28	-	-	0.95	0.90 to 0.98
Cyst volume	Longest length	31	-	-	0.91	0.84 to 0.95	30	-	-	0.83	0.68 to 0.92	30	-	-	0.83	0.67 to 0.91
	Widest width	31	-	-	0.83	0.72 to 0.91	31	-	-	0.95	0.91 to 0.98	32	-	-	0.95	0.90 to 0.98
	Greatest depth	31	-	-	0.92	0.83 to 0.96	31	-	-	0.85	0.71 to 0.92	32	-	-	0.92	0.83 to 0.96
	Cyst volume	30	-	-	0.95	0.89 to 0.97	30	-	-	0.92	0.84 to 0.96	30	-	-	0.90	0.80 to 0.95
	Change in volume	---	-----	-----	-----	-----	---	-----	-----	-----	-----	32	0.81	0.81	-	0.66 to 0.96
Tubulation	Present/absent	31	0.89	0.63	-	0.42 to 0.83	30	0.90	0.00	-	0.0 to 0.0	32	0.88	0.46	-	0.04 to 0.88
	Bone expansion	---	-----	-----	-----	-----	32	0.88	0.30	-	-0.16 to 0.76	32	0.88	0.45	-	0.027 to 0.87
	Widest width of tub.	22	-	-	0.97	0.93 to 0.99	27	-	-	0.99	0.97 to 0.99	26	-	-	0.98	0.96 to 0.99
	Expected bone width	22	-	-	0.85	0.73 to 0.93	27	-	-	0.97	0.93 to 0.98	26	-	-	0.94	0.87 to 0.97
Scalloping	Present/absent	32	0.69	0.30	-	0.15 to 0.45	32	0.81	-0.055	-	-0.15 to 0.038	32	0.81	0.42	-	0.077 to 0.76
	Percent of border	14	-	-	0.37	0.037 to 0.70	26	-	-	0.70	0.43 to 0.85	22	-	-	0.67	0.36 to 0.85
	Cyst index A	31	-	-	0.14	-0.062 to 0.39	30	-	-	0.88	0.77 to 0.94	29	-	-	0.85	0.70 to 0.92
Cyst index	Cyst index B	30	-	-	0.31	-0.05 to 0.60	30	-	-	0.86	0.72 to 0.93	29	-	-	0.84	0.68 to 0.92
	Presence of fracture	32	0.81	0.37	-	0.21 to 0.53	31	0.90	0.71	-	0.42 to 1.0	32	0.81	0.55	-	0.24 to 0.85
Cyst grade	Cyst grade	32	0.47	0.27	-	0.15 to 0.39	32	0.66	0.57	-	0.37 to 0.77	32	0.59	0.47	-	0.27 to 0.67
	Bone length	20	-	-	1.00	1.0 to 1.0	19	-	-	1.00	1.0 to 1.0	18	-	-	1.00	1.0 to 1.0
Bone measurements	Diameter of diaphysis	---	-----	-----	-----	-----	32	-	-	0.99	0.97 to 0.99	31	-	-	0.98	0.95 to 0.99
	Width of nearest physis to cyst	30	-	-	-0.61	-0.79 to -0.32	32	-	-	0.89	0.79 to 0.95	32	-	-	0.99	0.99 to 1.0

Single hyphens indicate value not calculated due to data classification as continuous or categorical. Dashed lines are indicative of a parameter that was irrelevant to that particular stage of reading
n, sample size; CI, confidence interval

Diagnostic factors, such as the cyst's location,^{4,7} number of loculations,^{7,8,37} the presence of a cortical rim,^{7,29,38} and evidence of periosteal reaction^{30,39} have been used to identify SBCs. The inter-rater reliability of the number of loculations and periosteal reaction in this study were similar or better compared with studies on other benign bone lesions such as aneurysmal bone cysts³⁰ or malignant bone lesions, and are useful parameters.³¹⁻³³ The criteria for the cortical rim demonstrated consistently poor reliability in this study due to the ambiguity of the cyst walls. Comparable reliability values for cortical rim in literature have varied depending on whether the study evaluated cortical rim thickening or destruction.³¹⁻³³ The lack of a consistent method of assessment and the variability in reliability indicates this parameter as an unreliable factor, therefore, we recommend against its use to assess SBCs.

Prognostic information such as scalloping of the cyst wall and tubulation⁴⁰ (widening) have been associated with non-healing and cyst recurrence. Additionally, active cysts^{1,29,37} or an increase in cyst volume^{7,41} have been associated with fracture. Prior studies of lesions such as enchondroma and chondrosarcoma have demonstrated high percent agreement,³³ with slight to fair kappa values.³¹ The reliability of tubulation, lesion activity and volume have not been previously investigated but demonstrated excellent values in this study and thus are useful parameters. The cyst index as proposed by Kaelin and McEwen¹³ is another prognostic tool that has been used to assess SBCs.^{1,8,10,42} Vasconcelos et al³⁴ reported poor inter-observer reliability most likely due to the use of trapezoidal shapes to approximate the cyst area for the equation. This can create unreliability due to the variable configuration of SBCs. To address this, one study used tracing software to determine the area and reported substantial reliability.³⁵ The current study created two modified cyst indices (A and B) based on a rectangular approximation. Both demonstrated excellent reliability in the second and third readings. However, as Cyst index B was more reliable in the first reading and also accounts for cyst depth, this is the preferred method.

Currently, the most commonly used measure of outcome in the form of cyst healing is the Neer Classification,¹⁴ and the modified version by Hashemi-Nejad and Cole.³ Reliability values in this study were consistently lower than previous studies reporting substantial or almost perfect reliability ($\kappa = 0.76$ to 0.88).^{35,36} This raises concerns that cyst healing may not be a reliable measure of treatment success. Identifying the presence of fracture had an excellent percentage agreement as previously reported.³³

The second objective of this study was to assess the reliability of reading radiographs individually *versus* in a chronological series. Other studies have reported that the serial chronological method of reading is both more sen-

sitive to change and has greater reliability,^{15,16} which could be due to overestimations caused by the bias of knowing that the series of images belongs to a single patient.^{15,43} This study did not find a significant difference between the methods. However, radiologists observed that a series provided more information for difficult assessments made after fractures or surgery, enabled more accurate land marking for measurements and assisted in cases where images within a series varied in quality and thus should be the preferred method.

There are many orthopaedic measurements that lack specific criteria, which can cause variability in clinical evaluations. Many studies of measurement variability assess reliability once and often conclude unreliability. This study was relatively unique in that after finding poor reliability values in the first reading, more specific criteria and a schematic were developed which enhanced the reliability of all but two parameters. As a result, variables that we recommend for use in the assessment of SBCs include all that were studied with the exception of the cortical rim. Future studies should aim to not only evaluate reliability but to also implement strategies to enhance reliability.

The main limitation for the current study is the generalizability of results. Reliability values and measurements were obtained in a study setting from two to three radiologists, all with long-term experience. However, radiologists did not receive a formalized tutorial session prior to readings, to mimic clinical practice and to minimize a false elevation in reliability due to a study setting. Second, this study only investigated variability due to the clinician.⁴⁴ Other sources of variance such as patient or procedure may affect the reliability and thus the reliability found in this study should be considered a slight overestimate of the true reliability.⁴⁴ Finally, this study evaluated the radiographic assessment of SBCs and, therefore, the results cannot necessarily be extrapolated to other benign bone lesions.

In conclusion, the majority of the radiographic criteria developed for the diagnosis, prognosis and outcome of SBC were found to be reliable and improved with explicit criteria and a visual schematic. Inter-rater reliability did not differ between single and serial-chronological readings.

Received 09 August 2018; accepted after revision 14 January 2019.

COMPLIANCE WITH ETHICAL STANDARDS

FUNDING STATEMENT

This study was funded by the Canadian Institutes of Health Research (CIHR, FRN 130272).

OA LICENCE TEXT

This article is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 International (CC BY-NC 4.0) licence (<https://creativecommons.org/>

licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed.

ETHICAL STATEMENT

Ethical approval: Approval for this study was granted by our institutional research ethics board. This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent: Not required for this type of study.

ICMJE CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflict of interest.

ACKNOWLEDGEMENTS

We would like to thank Peggy Law for her help on this project.

AUTHOR CONTRIBUTIONS

All authors: study design, data analyses and writing of the manuscript.

SC: data collection, all of the above.

REFERENCES

- Tey IK, Mahadev A, Lim KB, Lee EH, Nathan SS.** Active unicameral bone cysts in the upper limb are at greater risk of fracture. *J Orthop Surg (Hong Kong)* 2009;17:157-160.
- Docquier PL, Delloye C.** Treatment of simple bone cysts with aspiration and a single bone marrow injection. *J Pediatr Orthop* 2003;23:766-773.
- Hashemi-Nejad A, Cole WG.** Incomplete healing of simple bone cysts after steroid injections. *J Bone Joint Surg [Br]* 1997;79-B:727-730.
- Jaffe HL, Lichtenstein L.** Solitary unicameral bone cyst, with emphasis on the roentgen picture, pathologic appearance and pathogenesis. *Arch Orthop Trauma Surg* 1942;44:1001-1025.
- Azouz EM, Karamitsos C, Reed MH, et al.** Types and complications of femoral neck fractures in children. *Pediatr Radiol* 1993;23:415-420.
- Neer CS II, Francis KC, Johnston AD, Kiernan HA Jr.** Current concepts on the treatment of solitary unicameral bone cyst. *Clin Orthop Relat Res* 1973;97:40-51.
- Capanna R, Dal Monte A, Gitelis S, Campanacci M.** The natural history of unicameral bone cyst after steroid injection. *Clin Orthop Relat Res* 1982;166:204-211.
- Teoh KH, Watts AC, Chee Y-H, Reid R, Porter DE.** Predictive factors for recurrence of simple bone cyst of the proximal humerus. *J Orthop Surg (Hong Kong)* 2010;18:215-219.
- Hagmann S, Eichhorn F, Moradi B, et al.** Mid- and long-term clinical results of surgical therapy in unicameral bone cysts. *BMC Musculoskelet Disord* 2011;12:281.
- Norman-Taylor FH, Hashemi-Nejad A, Gillingham BL, Stevens D, Cole WG.** Risk of refracture through unicameral bone cysts of the proximal femur. *J Pediatr Orthop* 2002;22:249-254.
- Glowacki M, Ignys-O'Byrne A, Ignys I, Wroblewska K.** Limb shortening in the course of solitary bone cyst treatment—a comparative study. *Skeletal Radiol* 2011;40:173-179.
- Ahn JI, Park JS.** Pathological fractures secondary to unicameral bone cysts. *Int Orthop* 1994;18:20-22.
- Kaelin AJ, MacEwen GD.** Unicameral bone cysts. Natural history and the risk of fracture. *Int Orthop* 1989;13:275-282.
- Neer CS II, Francis KC, Marcove RC, Terz J, Carbonara PN.** Treatment of unicameral bone cyst. A follow-up study of one hundred seventy-five cases. *J Bone Joint Surg [Am]* 1966;48-A:731-745.
- van Der Heijde D, Boonen A, Boers M, Kostense P, van Der Linden S.** Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials. *Rheumatology (Oxford)* 1999;38:1213-1220.
- van Tuyl LH, van der Heijde D, Knol DL, Boers M.** Chronological reading of radiographs in rheumatoid arthritis increases efficiency and does not lead to bias. *Ann Rheum Dis* 2014;73:391-395.
- Wright JG, Yandow S, Donaldson S, Marley L; Simple Bone Cyst Trial Group.** A randomized clinical trial comparing intralesional bone marrow and steroid injections for simple bone cysts. *J Bone Joint Surg [Am]* 2008;90-A:722-730.
- Harris PA, Taylor R, Thielke R, et al.** Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377-381.
- Walter SD, Eliasziw M, Donner A.** Sample size and optimal designs for reliability studies. *Stat Med* 1998;17:101-110.
- Fleiss JL.** *The design and analysis of clinical experiments.* First ed. New York, NY: John Wiley and Sons, 1986.
- Fleiss JL, Cohen J.** The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973;33:613-619.
- Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.
- Feinstein AR, Cicchetti DV.** High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990;43:543-549.
- Lantz CA, Nebenzahl E.** Behavior and interpretation of the kappa statistic: resolution of the two paradoxes. *J Clin Epidemiol* 1996;49:431-434.
- Warrens MJ.** A formal proof of a paradox associated with Cohen's kappa. *J Classif* 2010;27:322-332.
- Shankar V, Bangdiwala SI.** Observer agreement paradoxes in 2x2 tables: comparison of agreement measures. *BMC Med Res Methodol* 2014;14:100.
- Sim J, Wright CC.** The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257-268.
- Nelson JC, Pepe MS.** Statistical description of interrater variability in ordinal ratings. *Stat Methods Med Res* 2000;9:475-496.
- Openheim WL, Galleno H.** Operative treatment versus steroid injection in the management of unicameral bone cysts. *J Pediatr Orthop* 1984;4:1-7.
- Mahnken AH, Nolte-Ernting CC, Wildberger JE, et al.** Aneurysmal bone cyst: value of MR imaging and conventional radiography. *Eur Radiol* 2003;13:1118-1124.
- Geirnaerdt MJ, Hermans J, Bloem JL, et al.** Usefulness of radiography in differentiating enchondroma from central grade 1 chondrosarcoma. *AJR Am J Roentgenol* 1997;169:1097-1104.
- Gourtsoyianni S, Hwang S, Panicek DM, et al.** Reproducibility and clinical correlations of post-treatment changes on CT of prostate cancer bone metastases treated with chemotherapy. *Br J Radiol* 2012;85:1243-1249.

33. **Gajewski DA, Burnette JB, Murphey MD, Temple HT.** Differentiating clinical and radiographic features of enchondroma and secondary chondrosarcoma in the foot. *Foot Ankle Int* 2006;27:240-244.
34. **Vasconcellos DA, Yandow SM, Grace AM, et al.** Cyst index: a nonpredictor of simple bone cyst fracture. *J Pediatr Orthop* 2007;27:307-310.
35. **Cha SM, Shin HD, Kim KC, Kang DH.** Flexible intramedullary nailing in simple bone cysts of the proximal humerus: prospective study for high-risk cases of pathologic fracture. *J Pediatr Orthop B* 2013;22:475-480.
36. **Cho HS, Seo SH, Park SH, et al.** Minimal invasive surgery for unicameral bone cyst using demineralized bone matrix: a case series. *BMC Musculoskelet Disord* 2012;13:134.
37. **Campanacci M, Capanna R, Picci P.** Unicameral and aneurysmal bone cysts. *Clin Orthop Relat Res* 1986;204:25-36.
38. **Chang CH, Stanton RP, Glutting J.** Unicameral bone cysts treated by injection of bone marrow or methylprednisolone. *J Bone Joint Surg [Br]* 2002;84-B:407-412.
39. **Ragsdale BD, Madewell JE, Sweet DE.** Radiologic and pathologic analysis of solitary bone lesions. Part II: periosteal reactions. *Radiol Clin North Am* 1981;19:749-783.
40. **Manaster BJ, May DA, Disler DG.** *Musculoskeletal imaging: The requisites*. 4th ed. Philadelphia: Elsevier Health Sciences, 2013.
41. **Glowacki M, Ignys-O'Byrne A, Ignys I, Mankowski P, Melzer P.** Evaluation of volume and solitary bone cyst remodeling using conventional radiological examination. *Skeletal Radiol* 2010;39:251-259.
42. **Ulici A, Balanescu R, Topor L, Barbu M.** The modern treatment of the simple bone cysts. *J Med Life* 2012;5:469-473.
43. **Ferrara R, Priolo F, Cammisa M, et al.** Clinical trials in rheumatoid arthritis: methodological suggestions for assessing radiographs arising from the GRISAR Study. Gruppo Reumatologi Italiani Studio Artrite Reumatoide. *Ann Rheum Dis* 1997;56:608-612.
44. **Wright JG, Feinstein AR.** Improving the reliability of orthopaedic measurements. *J Bone Joint Surg [Br]* 1992;74-B:287-291.