

Assessing the reliability of the modified Gartland classification system for extension-type supracondylar humerus fractures

T. L. Teo¹
E. K. Schaeffer¹
E. Habib¹
A. Cherukupalli¹
A. P. Cooper¹
A. Aroojis²
W. N. Sankar³
V. V. Upasani⁴
S. Carsen⁵
K. Mulpuri¹
C. Reilly¹

Abstract

Purpose The Gartland extension-type supracondylar humerus (SCH) fracture is the most common paediatric elbow fracture. Treatment options range from nonoperative treatment (taping or casting) to operative treatments (closed reduction and percutaneous pinning or open reduction). Classification variability between surgeons is a potential contributing factor to existing controversy over treatment options for type II SCH fractures. This study investigated levels of agreement in extension-type SCH fracture classification using the modified Gartland classification system.

Methods A retrospective review was conducted on 60 patients aged between two and 12 years who had sustained an extension-type SCH fracture and received operative or nonoperative treatment at a tertiary children's hospital. Baseline radiographs were provided, and surgeons were asked to classify the fractures as type I, IIA, IIB or III according to the mod-

ified Gartland classification. Respondents were then asked to complete a second round of classifications using reshuffled radiographs. Weighted kappa values were calculated to assess interobserver and intraobserver levels of agreement.

Results In all, 21 paediatric orthopaedic surgeons responded to the survey and 15 completed a second round of ratings. Interobserver agreement for classification based on the Gartland criteria between surgeons was substantial with a kappa of 0.679 (95% confidence interval (CI) 0.501 to 0.873). Intraobserver agreement was substantial with a kappa of 0.796 (95% CI 0.628 to 0.864).

Conclusion Radiographic classification of extension-type SCH fractures demonstrated substantial agreement both between and within surgeon raters. Therefore, classification variability may not be a major contributing factor to the treatment controversy for type II SCH fractures and treatment variability may be due to differences in surgeon preferences.

Level of Evidence: III

Cite this article: Teo TL, Schaeffer EK, Habib E, Cherukupalli A, Cooper AP, Aroojis A, Sankar WN, Upasani VV, Carsen S, Mulpuri K, Reilly C. Assessing the reliability of the modified Gartland classification system for extension-type supracondylar humerus fractures. *J Child Orthop* 2019;13:569-574. DOI: 10.1302/1863-2548.13.190005

Keywords: supracondylar humerus fracture; paediatrics; elbow fracture; Gartland classification; reliability

Introduction

The Gartland extension-type supracondylar humerus fracture is the most common paediatric elbow fracture.¹ Depending on fracture classification, treatment options range from nonoperative, such as closed reduction and/or tape, splint or cast immobilization, to operative, such as closed reduction and percutaneous pinning or open reduction and pinning.² There is generally little controversy over whether nonoperative *versus* operative management is recommended for Gartland types I and III fractures. However, controversy still exists around the world for type II fracture treatment; many surgeons prefer operative treatment while some surgeons choose to treat type II fractures nonoperatively. Despite an increased

¹Department of Orthopaedic Surgery, British Columbia Children's Hospital, Vancouver, British Columbia, Canada

²Department of Paediatric Orthopaedics, B.J. Wadia Hospital for Children, Mumbai, India

³Children's Hospital of Philadelphia, Philadelphia, Philadelphia, USA

⁴Rady Children's Hospital San Diego, University of California San Diego, San Diego, California, USA

⁵Division of Orthopedic Surgery, University of Ottawa, Ottawa, Ontario, Canada

Correspondence should be sent to K. Mulpuri, Department of Orthopaedic Surgery, BC Children's Hospital, 1D.66-4480 Oak Street, Vancouver, British Columbia, V6H 3V4, Canada.
E-mail: kmulpuri@cw.bc.ca

trend towards operative treatment for type II fractures, the current evidence does not completely support the superiority of operative over nonoperative methods.³

A potential contributing factor to this controversy is variability in classification of fractures by different orthopaedic surgeons. Literature generally supports the use of the modified Gartland classification system; however, the classification of certain subtypes of supracondylar fractures can be quite variable between surgeons, including the following treatment regimen.⁴ Differences in classification can mean the difference between operative and nonoperative treatment for a patient. For example, a patient with a borderline type I/IIA fracture may get treated with a splint if classified as type I or surgery if classified as type II if they went to a surgeon who chooses to treat type IIA fractures operatively. Similarly, a patient with a borderline type IIA/IIB fracture may get treated with casting if classified as type IIA or surgery if classified as type IIB if they went to a surgeon who chooses to treat type IIA fractures nonoperatively.

It is necessary to determine whether differences in treatment preferences for type II fractures between surgeons are due to a true difference in practice patterns or due to surgeons classifying the same fractures differently. If a true difference in practice patterns exists, then there is a need for further research to compare outcomes between nonoperative and operative management in order to standardize patient care. If, however, the differences in treatment preferences are due to different patterns of classification between surgeons, then it may call into question the utility of the modified Gartland classification system in determining management. In this case, further exploration into the factors influencing individual surgeon decision-making may allow for the development of a more reliable diagnostic classification.

To date, there has been no classification reliability study performed between surgeons across countries. The purpose of this study was to investigate the levels of agreement between surgeons from Canada, the United States (USA), Australia, the United Kingdom (UK), and India in the classification of extension-type supracondylar humerus fractures using the modified Gartland classification system.

Materials and methods

After receiving institutional Research Ethics Board approval at the University of British Columbia, a retrospective radiographic and chart review was conducted on patients aged two to 12 years who had sustained an extension-type supracondylar humerus fracture between January 2005 and December 2016, received either operative or nonoperative treatment at a tertiary paediatric hospital, and had adequate pre-reduction radiographs available. Radiograph

adequacy was defined by a true anteroposterior (AP) view on AP radiographs with orthogonal visualization of the distal humerus and clear delineation of the hourglass sign and capitellum on lateral radiographs.

A total of 60 patients were selected for inclusion after review. These patients had received a Gartland classification diagnosis by one of seven surgeons from a single institution: ten were diagnosed with type I fractures, 25 with type II fractures and 25 with type III fractures. Many of the cases were chosen because they straddled the borderline between two categories, to reflect the difficulty in decision-making often seen in clinical practice. Each patient had an adequate set of baseline AP and lateral plain elbow radiographs as determined by the senior author (CR); these were de-identified and compiled into surveys administered through Research Electronic Data Capture (REDCap) software (Vanderbilt University, Nashville, Tennessee).

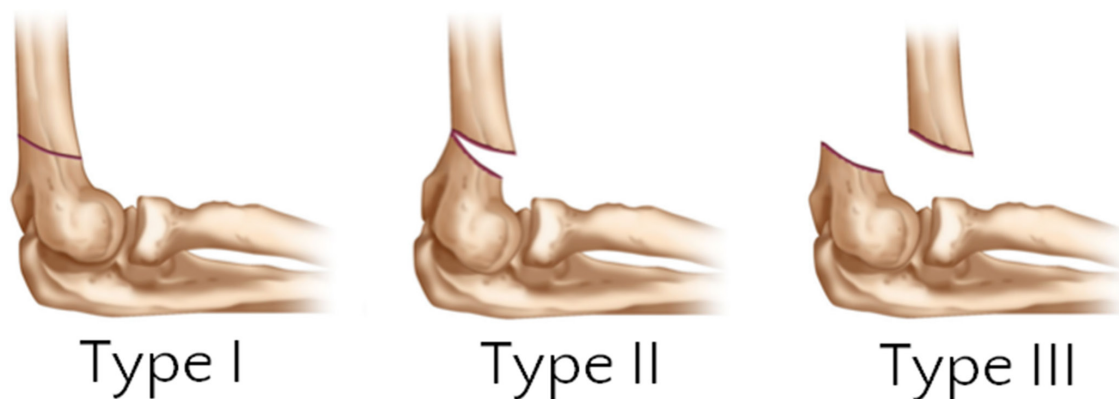
Invitations to participate in the survey were sent out to fellowship-trained paediatric orthopaedic surgeons practising in tertiary care hospitals around the world. This study was conducted in two rounds. For the first round, surgeons were provided with a brief pictorial and table summary of the Wilkins-modified⁵ Gartland classification system, along with the compiled radiographs (Fig. 1). Each surgeon was blinded to patient treatment and original diagnosis and asked to classify the fractures as type I, IIA, IIB or III according to the system provided. In order to assess intraobserver agreement, the same radiographs were reshuffled and surgeons were asked to reclassify each set of radiographs using the same classification system following a two-week interval.

Computer-generated weighted kappa statistics using pairwise comparisons were calculated along with 95% confidence intervals (CI) to assess interobserver and intraobserver levels of agreement. The weighted kappa coefficients used (Table 1) took into consideration the varying levels of clinical importance for disagreement between each classification (e.g. a disagreement between type I and IIA would be less significant than one between type I and III) as determined by two orthopaedic surgeons (CR, KM). The kappa values were interpreted using the Landis and Koch guidelines⁶ outlined as follows: values < 0.00 indicate poor agreement; 0.00 to 0.20 slight agreement; 0.21 to 0.40 fair agreement; 0.41 to 0.60 moderate agreement; 0.61 to 0.80 substantial agreement; and 0.81 to 1.00 excellent or almost perfect agreement.

Results

Participant demographics

A total of 31 surgeons comprised of members of the Canadian Pediatric Orthopaedic Group and surgeons known to



Type I	Non-displaced
Type II A	Intact posterior cortex, hinged in extension. No rotation or translation.
Type II B	Intact posterior cortex, hinged in extension, with some degree of rotational displacement or translation.
Type III	Complete displacement.

Fig. 1 Summary of Wilkins-modified Gartland classification system⁵ provided to survey respondents. Figure 1 is being reprinted with permission and licensing from Wolters Kluwer Health, Inc. Journal content for figure re-use: Authors: Timothy Alton, Shawn Werner, and Albert Gee; Article Title: Classifications In Brief: The Gartland Classification of Supracondylar Humerus Fractures; Journal: Clinical Orthopaedics and Related Research; Volume 473; Issue 2; Pages 738-741; URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4294919/>

Table 1 Relative weights assigned to kappas for disagreement between observers

Disagreement between classifications	Weight
Type I vs IIA	1
Type I vs IIB	4
Type I vs III	6
Type IIA vs IIB	3
Type IIA vs III	4
Type IIB vs III	1

one of the study authors (KM) were invited to participate. In all, 21 surgeons (14 from Canada, three from the USA, two from Australia, one from the UK, one from India) representing 17 tertiary care hospitals responded to the first round of surveys. Of these respondents, four were from the same institution. The mean length of practice across all respondents was ten years (1 to 26). The majority of the respondents treat approximately 41 to 60 supracondylar humerus fractures a year. None of the respondents treat < 20 a year. After a two-week period, 15 of the original survey respondents completed the second round of surveys.

Table 2 Weighted interobserver and intraobserver kappas (κ)

Patient cases, n	Combined weighted interobserver κ (95% CI) (n = 21)	Combined weighted intraobserver κ (95% CI) (n=15)
60	0.679 (0.468 to 0.889)	0.796 (0.628 to 0.964)

CI, confidence interval

Interobserver level of agreement

The weighted interobserver agreement for classification based on the modified Gartland criteria across all respondents was substantial (Table 2). Levels of agreement were similarly substantial for the Canadian and non-Canadian groups, with weighted kappa values of 0.687 (95% CI 0.501 to 0.873) and 0.663 (95% CI 0.436 to 0.891), respectively. Across the four respondents from the same institution, the level of agreement was substantial with a weighted kappa of 0.746 (95% CI 0.654 to 0.839). The most common source of disagreement was between the type IIA and IIB classifications.

Table 3 Comparison with other studies evaluating the reliability of the Gartland classification system

Study	Number of patient cases	Number of raters	Raters' country of practice	Classification system	Overall interobserver κ	Interpretation	Overall intraobserver κ	Interpretation
Barton et al (2001) ⁴	50	5	United States	Modified Gartland	0.74	Substantial	0.81 to 0.84	Excellent or almost perfect
Heal et al (2007) ⁷	50	4	United Kingdom	Modified Gartland	0.54	Moderate	0.68 to 0.83	Substantial to excellent or almost perfect
Mallo et al (2010) ²	75	4	United States	Standard Gartland	0.521	Moderate	0.723	Substantial
Rocha et al (2015) ⁸	50	3	Brazil	Standard Gartland	0.756	Substantial	0.719 to 0.859	Substantial
Leung et al (2018) ⁹	200	5	United States	Modified Gartland	0.475	Moderate	0.777	Substantial
Present study	60	21	Canada, United States, United Kingdom, Australia, India	Modified Gartland	0.679	Substantial	0.796	Substantial

κ , kappa

Intraobserver level of agreement

The weighted intraobserver agreement for classification based on the modified Gartland criteria for all respondents was substantial (Table 2). Individual intraobserver agreements for each of the 15 respondents who completed both surveys ranged from 0.554 (moderate) to 0.898 (excellent or almost perfect).

Discussion

The results of this study demonstrate a substantial level of interobserver and intraobserver agreement in the radiographic classification of extension-type supracondylar humerus fractures at baseline. The levels of agreement are substantial enough to suggest that classification variability is not a major contributing factor to variability in treatment between surgeons for type II supracondylar fractures. However, as levels of agreement were not perfect, we acknowledge that there are still a proportion of cases for which treatment decisions may vary based on the way individual surgeons classify them.

The incidence of paediatric supracondylar fractures has increased over the last decade, potentially due to the greater prevalence of high-energy recreational activities that children participate in in recent years.³ In an effort to standardize the management of these fractures, the Gartland classification system and treatment algorithm was created in 1959 and continues to form the basis of the American Academy of Orthopaedic Surgeons' treatment recommendations for paediatric supracondylar fractures.¹ Despite this, controversy still exists over the necessity of operative treatment in the management of Gartland type IIA fractures. While this may reflect true differences in surgeon preferences, it may also potentially be the result of classification differences, where a fracture may be classified as type I and thus treated nonoperatively by one surgeon and classified as IIA or IIB by another and treated operatively. Consequently, we need to ascertain

that fractures are being classified consistently in order to compare treatment outcomes for nonoperative and operative methods within a single fracture type. This would not only allow for confidence when interpreting results of potential multi-centre prospective study, but also allow for meaningful comparisons across multiple studies with the knowledge that surgeons are discussing the same kinds of fractures.

Previous studies have investigated the reliability of the Gartland classification system (Table 3). Barton et al⁴ performed a study with five physicians of varying levels of training and areas of expertise – a junior orthopaedic resident, a senior orthopaedic resident, a paediatric orthopaedic fellow, an attending-level paediatric orthopaedic surgeon and an attending-level paediatric orthopaedic radiologist. The study found substantial interobserver agreement and excellent intraobserver agreement. A subsequent study involving four orthopaedic surgeons based in the UK found moderate interobserver agreement overall, poor interobserver agreement for type I fractures, fair to moderate interobserver agreement for type II fractures, substantial to excellent interobserver agreement for type III fractures and substantial to excellent intraobserver agreement overall.⁷ A later study in 2010 involving four fellowship-trained paediatric orthopaedic surgeons based in New York confirmed moderate interobserver agreement overall, moderate interobserver agreement for type I fractures, moderate interobserver agreement for type II fractures, substantial interobserver agreement for type III fractures and substantial intraobserver agreement overall.² Both of the latter studies suggested that injury mechanism, soft-tissue status and neurovascular compromise are considered by surgeons in addition to the Gartland classification in making treatment decisions. A more recent study from Brazil between three paediatric orthopaedic surgeons found substantial interobserver agreement overall, excellent interobserver agreement for type I fractures, moderate interobserver agreement for type II fractures, substantial interobserver agreement for type III

fractures and excellent intraobserver agreement overall.⁸ Finally, a study by Leung et al⁹ between five USA-based surgeons found moderate interobserver agreement overall and substantial intraobserver agreement overall. They questioned the utility of the Wilkins-modified classification system in guiding clinical decision-making, particularly for type II fractures where the reliability of classification between surgeons is low. They also found a much greater level of agreement when asked to classify fractures as requiring operative or nonoperative treatment instead.

Two observations may be drawn from these studies. Firstly, while the levels of interobserver agreement for types I and III fractures and general intraobserver agreement tends to be high, interobserver agreement for type II fractures tend to be only as good as fair to moderate. We chose not to perform subgroup analyses for individual classification types because it would have required using the original diagnosis as the gold standard for categorization, which is subjective based on individual surgeon variation. This was also evident in a study by Hyman et al¹⁰ where classification of staging in patients with Legg-Calvé-Perthes disease was limited by a lack of a benchmark for comparison. They also considered the use of a senior author as a standard of comparison; however, such an approach was limited due to the inability to determine the accuracy of the author's rating. Nevertheless, we attempted to account for this by using weighted kappa coefficients in our calculations to reflect the varying degrees of clinical importance for discrepancies between various fracture classifications. We also note that our observed frequency of disagreement was indeed highest between type IIA and IIB classifications. Secondly, none of these studies has examined agreement between surgeons practising across different countries and thus their results may not accurately reflect differences between surgeons around the world. The present study attempted to improve upon the generalizability of the findings by examining agreement between a significantly larger group of surgeon respondents representing multiple tertiary paediatric centres around the world.

There are limitations to this study. The surgeons invited to respond to this survey were all known to one of the study's authors (KM) and many of them had completed their one-year paediatric orthopaedic fellowship at the same institution. This could have potentially increased the level of agreement above the true value since the surgeons' diagnostic patterns would have been influenced by having been through similar training. Nevertheless, these surgeons also performed their residencies at different institutions and have practiced for a number of years at their current institutions, where their practice patterns would inevitably be influenced differently. Another limitation to this study is that a large proportion of the survey respondents were Canadian, which challenges the generalizability of the study results to countries outside of Canada.

Most of the survey respondents practise in Western countries, with only one practising in an Asian country, so the study was not as internationally representative of an assessment as intended. However, this is the first study to assess classification reliability using respondents from more than one institution or country. In order to assess the extent of difference that having a large proportion of Canadian respondents would cause, we examined the levels of agreement when Canadians and non-Canadians were grouped separately and found that, within the limitations of our sample size, levels of agreement were similar between both groups, although a larger number of respondents would be needed in order to test this fully. Furthermore, while the level of agreement between the four surgeons from the same institution appears to be slightly better than the overall level of agreement, both kappa values still fall within the substantial agreement category, suggesting that neither shared institution nor shared country of practice have a large influence on the reliability in classification between surgeons. Another limitation to the study is that the weighted kappa coefficients that were used to account for the varying levels of clinical importance for disagreement between each classification were determined by only two orthopaedic surgeons. The weighted coefficients were used in an attempt to numerically account for the difference in clinical significance of disagreement between different Gartland classifications, where a disagreement between type IIA and IIB fractures is clinically more significant than a disagreement between type I and IIA fractures. This was important since we were calculating a quantitative kappa value for agreement using categories between which the clinical significance of disagreement is not equal. However, we acknowledge that having more than two surgeons determine the respective weighting used would have increased in the generalizability of our results. A further limitation was that we did not have representation from physicians or surgeons practising at community-based hospitals since all of the respondents were paediatric orthopaedic surgeons practising at tertiary referral hospitals. This might prevent the generalizability of the findings to non-Western countries and non-tertiary hospital-based practices, since training, methods of diagnosis and treatment may potentially differ in those areas. Finally, no residents or fellows were included in this survey, so the results may be limited to fellowship-trained attending-level paediatric orthopaedic surgeons.

The current orthopaedic literature is often limited by a lack of high-quality, scientific data, particularly in the treatment of paediatric supracondylar humerus fractures, where levels of evidence are limited by methodological shortcomings inherent in retrospective study designs.¹¹ There is a clear need for larger, prospective controlled trials to more effectively compare outcomes from different treatment modalities and these trials would ideally include

multiple centres so as to increase the generalizability of findings. However, in order to perform such cross-collaborations, we need to be certain that surgeons from different centres are classifying supracondylar humerus fractures in the same way. Unfortunately, while there should be a central adjudication system to standardize classification in an ideal trial, this is often limited by availability of resources.

In conclusion, this study suggests that levels of agreement between orthopaedic surgeons are substantial enough to allow for the meaningful comparison of extension-type paediatric supracondylar humerus fractures across the practices of attending-level orthopaedic surgeons practicing in Western countries. Classification variability does not seem to be a major contributing factor to the treatment controversy for type II supracondylar humerus fractures. Further research is needed to compare patient outcomes between nonoperative and operative treatment for these fractures, so as to establish consensus and a standardized treatment protocol for optimal patient care across centres.

Received 23 April 2019; accepted after revision 4 November 2019.

COMPLIANCE WITH ETHICAL STANDARDS

FUNDING STATEMENT

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

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: This study was reviewed and received institutional approval from the Research Ethics Board at the University of British Columbia (H16-02015).

Informed consent: This research involved surgeons as participants in a survey; therefore, consent was implied through survey completion.

ICMJE CONFLICT OF INTEREST STATEMENT

EKS reports grants from Pediatric Orthopaedic Society of North America, during the conduct of the study.

AnC reports non-financial support from Pega Medical, academic consultation fees from Vilex, grants from Vilex, all outside the submitted work.

VVU reports personal fees from OrthoPediatrics, personal fees from DePuy Synthes Spine, personal fees from Nuvasive, personal fees from Wolters-Kluwer Health, all outside the submitted work.

KM reports grants from International Hip Dysplasia Institute during the conduct of the study; grants from Depuy, Johnson & Johnson, grants from Pega Medical, grants from Allergan, grants from I'm a HIPpy Foundation, all outside the submitted work. In addition, KM has a Pega Medical patent pending.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Jeffrey Bone for his valuable assistance with the data analysis for this study. We also wish to thank Drs. Alex Aarvold, Benjamin Shore,

Debra Bartley, David Bade, Farhad Moola, Caroline Forsythe, Heather Jackman, Elaine Joughin, Leo Donnan, Lise Leveille, Mark Camp, Patricia Larouche, Ron El-Hawary and Timothy Carey for their substantial contribution of time as surgeon respondents for this study.

AUTHOR CONTRIBUTIONS

All authors listed have contributed sufficiently to the manuscript to be included as authors, and all those who are qualified to be authors are listed in the author byline.

TLT: Developed the study, Analyzed the data, Wrote the manuscript.

EKS: Contributed to the study development, Edited the manuscript.

EH: Made intellectual contributions, Edited the manuscript.

AC: Made intellectual contributions, Edited the manuscript.

APC: Made intellectual contributions, Edited the manuscript.

AA: Made intellectual contributions, Edited the manuscript.

WNS: Made intellectual contributions, Edited the manuscript.

VVU: Made intellectual contributions, Edited the manuscript.

SC: Made intellectual contributions, Edited the manuscript.

KM: Oversaw study completion and data interpretation, Advised on the writing of the manuscript.

CR: Oversaw study completion and data interpretation, Advised on the writing of the manuscript.

REFERENCES

1. **Alton TB, Werner SE, Gee AO.** Classifications in brief: the Gartland classification of supracondylar humerus fractures. *Clin Orthop Relat Res* 2015;473:738-741.
2. **Mallo G, Stanat SJC, Gaffney J.** Use of the Gartland classification system for treatment of pediatric supracondylar humerus fractures. *Orthopedics* 2010;33:19-22.
3. **Sinikumpu JJ, Pokka T, Sirviö M, Serlo W.** Gartland type II supracondylar humerus fractures, their operative treatment and lateral pinning are increasing: a population-based epidemiologic study of extension-type supracondylar humerus fractures in children. *Eur J Pediatr Surg* 2017;27:455-461.
4. **Barton KL, Kaminsky CK, Green DW, et al.** Reliability of a modified Gartland classification of supracondylar humerus fractures. *J Pediatr Orthop* 2001;21:27-30.
5. **Wilkins KE, Beaty JH, Chambers HG, et al.** Fractures and dislocations of the elbow region. Rockwood CA, Wilkins KE, Beaty J, eds. *Fractures in children*. Vol 4. Philadelphia, PA: Lippincott Williams and Wilkins, 1996:653-904.
6. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.
7. **Heal J, Bould M, Livingstone J, Blewitt N, Blom AW.** Reproducibility of the Gartland classification for supracondylar humeral fractures in children. *J Orthop Surg (Hong Kong)* 2007;15:12-14.
8. **Rocha ITS, Faria AS, Filho CF, Rocha MA.** Reproducibility of the AO/ASIF and Gartland classifications for supracondylar fractures of the humerus in children. *Rev Bras Ortop* 2015;50:266-269.
9. **Leung S, Paryavi E, Herman MJ, Sponseller PD, Abzug JM.** Does the modified Gartland classification clarify decision making? *J Pediatr Orthop* 2018;38:22-26.
10. **Hyman JE, Trupia EP, Wright ML, et al.** Interobserver and intraobserver reliability of the modified Waldenström classification system for staging of Legg-Calvé-Perthes disease. *J Bone Joint Surg [Am]* 2015;97-A:643-650.
11. **Howard A, Mulpuri K, Abel MF, et al.** The treatment of pediatric supracondylar humerus fractures. *J Am Acad Orthop Surg* 2012;20:320-327.