DOJ

10.5005/jp-journals-10017-1063

CLINICAL STUDY

Independent Analysis of the Dorr Classification of Proximal Femoral Morphology: A Reliability Study

[1]**Jason M Jennings** MD, DPT, [2]**Ramon A Ruberte Thiele** MD, MS, [3]**Evgeny Krynetskiy** MD, [4]**Samuel S Wellman** MD
[5]**David E Attarian** MD, FACS, [6]**Michael P Bolognesi** MD

ABSTRACT

The purpose of this study was to assess the inter- and intraobserver reliability through visual perception of a previously described classification system of proximal femoral morphology. The interobserver reliability ranged from slight to moderate across testing. Experience appears to play a role as the intratester reliability did not differ with fellowship trained attending physicians, but differences were seen with junior- and senior-level residents. The diversity of the proximal femoral morphology likely represents a continuum rather than three distinct shapes. This may imply that for many proximal femoral canals, characterization of a single type may not be possible. Further research is warranted to determine the clinical significance of these findings.

Keywords: Bone quality, Proximal femoral morphology, Total hip arthroplasty.

Jennings JM, Thiele RAR, Krynetskiy E, Wellman SS, Attarian DE, Bolognesi MP. Independent Analysis of the Dorr Classification of Proximal Femoral Morphology: A Reliability Study. The Duke Orthop J 2016;6(1):12-16.

Source of support: Nil

Conflict of interest: None

INTRODUCTION

Femoral component position is a predictor of long-term outcome in total hip arthroplasty (THA). The diversity in proximal femoral morphology has been well documented, and plays a critical role in prosthesis selection and fit, particularly with an uncemented femoral stem.[1-3]

A commonly used X-ray classification system of the proximal femur was proposed by Dorr et al.[1] This group described three distinct patterns of shape and bone structure of the proximal femur, labeled types A to C. The differences in the qualitative assessment of the proximal femur were defined and validated by shape and cortical thickness with specific radiographic measurements. It was suggested that such classification of the proximal femoral morphology was reliable since the interobserver variation was less than 20% at first review and improved to less than 5% at the second viewing.[1] However, this data relied on observations performed by two physicians.

To our knowledge, there has not been a study comparing the reliability of radiographic analyses using the methods previously described. The purpose of this study was to investigate the reliability of the Dorr classification system in defining proximal femoral morphology with a subgroup analysis based on physician level of training.

MATERIALS AND METHODS

This study was approved by the Institutional Review Board at our institution. Radiographs adjusted for magnification included anterior–posterior and frog-leg lateral views of the proximal 1/3 to 1/2 of the femur. All patients had evidence of primary osteoarthritis. Two separate investigators, who were not involved with the reliability portion of the study, measured the cortical index (CI) on the anterior–posterior and lateral views as previously described to qualitatively determine the types (A to C) of the proximal femur morphology (Figs 1A and B).[1] Only radiographs showing a similar CI measurement between the two investigators and showing a CI falling within the range of previously described types were included in this study.[1] The CI was used as it was the only radiographic measure reported to be significantly different for each type.[1] A total of 30 radiographs with an equal distribution of morphology types for analysis were utilized.

The physicians participating in the study were separated into three groups based on their level of training. Group I consisted of fellowship adult reconstructive attending physicians and fellows, group II senior-level residents, and group III junior residents. Prior to beginning the study, all participants reviewed the original article by Dorr et al[1] defining the classification system of proximal femoral morphology. In addition, immediately prior to each round of evaluation, a brief presentation was given with examples of each type of proximal femoral morphology. The selected radiographs were converted to a PowerPoint presentation, cropped appropriately to

[1-6]Orthopaedic surgeon

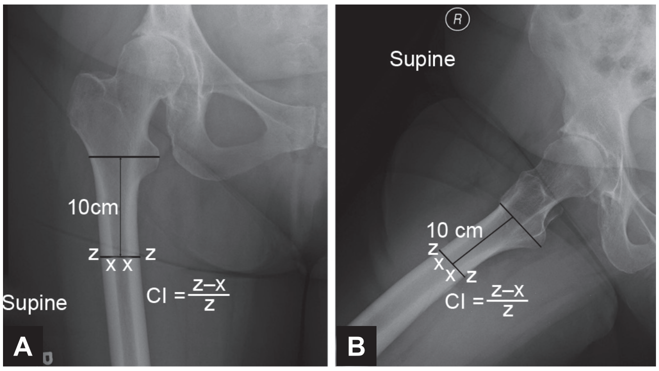
[1]Department of Orthopedic Surgery, Colorado Joint Replacement Porter Adventist Hospital, Denver, Colorado, USA

[2]Department of Orthopedic Surgery, University of Cincinnati Academic Health Center, Durham, North Carolina, USA

[3]Joint Implant Surgeons of Florida, Fort Myers, Florida, USA

[4-6]Department of Orthopedic Surgery, Duke University Medical Center, Durham, North Carolina, USA

Corresponding Author: Jason M Jennings, Orthopaedic Surgeon, 2535 South Downing Street, Denver, Colorado-80210 USA, e-mail: jasonjennings@centura.org



Figs 1A and B: Cortical index on the anterior–posterior and lateral views to qualitatively determine the proximal femoral morphology types (A to C)

allow for blinding, and viewed by participants in a conference room. The results were collected and the same test, with images in a new random order, was performed 2 weeks later to assess intratester reliability and changes in intertester reliability. Participants were blinded to the distribution of morphology types included in the test.

STATISTICAL METHODS

Fleiss's kappa was used to measure the inter-rater reliability across all readers and within types (A to C) and physician reader subgroups. The kappa statistic expresses the extent to which the observed amount of agreement among raters exceeds what would be expected if all raters made their ratings completely at random. Kappa values with 95% confidence intervals are presented.

Bootstrap methods were used to compare the correlated measures in rounds 1 and 2. A p-value < 0.05 was considered statistically significant. SAS version 9.2 was used for all analyses (SAS Institute, Inc.).

A commonly cited scale, while by no means universally accepted, is shown in Table 1.[4] Almost perfect agreement would equate to a kappa of 1, while agreement by chance would equate to 0.

Table 1: Kappa interpretation scale

0.0	0.20	0.40	0.60	0.80	1.00
Poor	Slight	Fair	Moderate	Substantial	Almost perfect

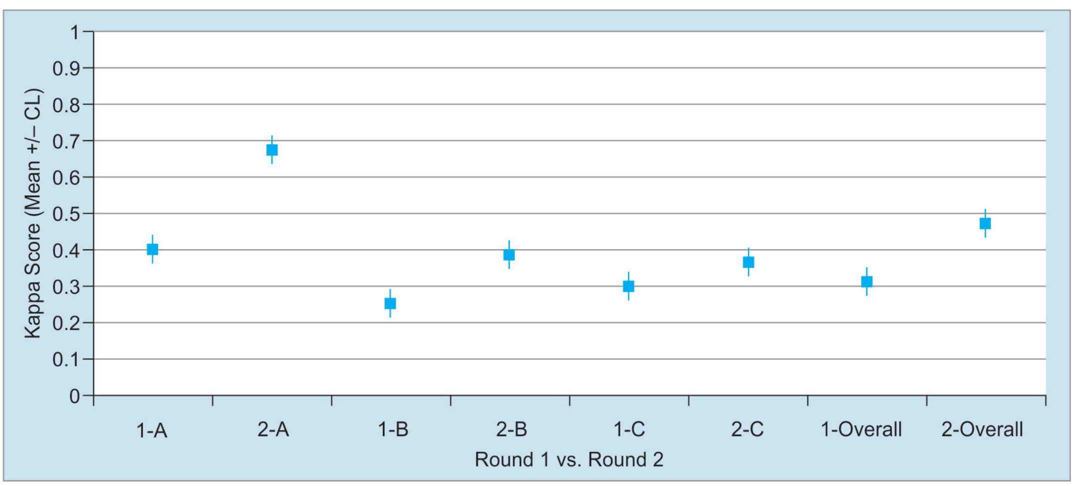
RESULTS

There were a total of 14 participants included in the study. Group I consisted of two adult reconstruction fellowship trained attending surgeons and one adult reconstructive fellow. Group II had six PGY 4 and 5 orthopedic surgery residents. Group III consisted of five PGY 2 and 3 orthopedic surgery residents.

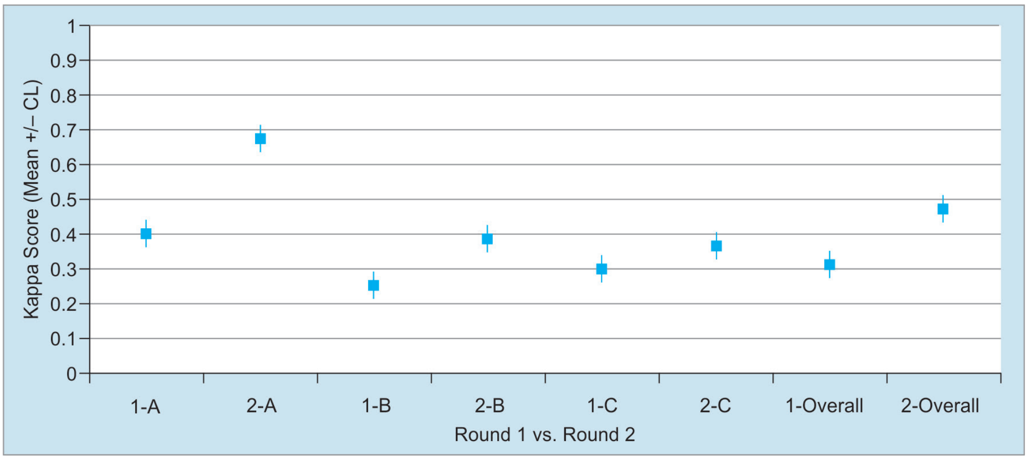
In round 1, the highest inter-rater agreement was obtained for type A (0.40) and the lowest for type B (0.25). In round 2, the highest interobserver agreement was obtained for type A (0.67), and the lowest for type C (0.36). Comparing rounds 1 to 2, significant improvement was seen in the identification of type A (0.40 *vs* 0.67), but not of types B (0.25 *vs* 0.39) and C (0.30 *vs* 0.36) (Graph 1). Overall, kappa scores for the identification of types significantly improved from rounds 1 to 2 (0.31 *vs* 0.47).

Group I obtained the highest interobserver agreement for type A (0.74) and the lowest for type B (0.54) in round 1. This subgroup obtained the highest interobserver agreement for type A (0.79) and the lowest for type C (0.32) in round 2. Comparing rounds 1 to 2, no significant improvement was seen in the identification of type A (0.74 *vs* 0.79), type B (0.55 *vs* 0.46), or type C (0.59 *vs* 0.32) (Graph 2). Overall, group I kappa scores for the identification of types did not significantly improve from rounds 1 to 2 (0.62 *vs* 0.55).

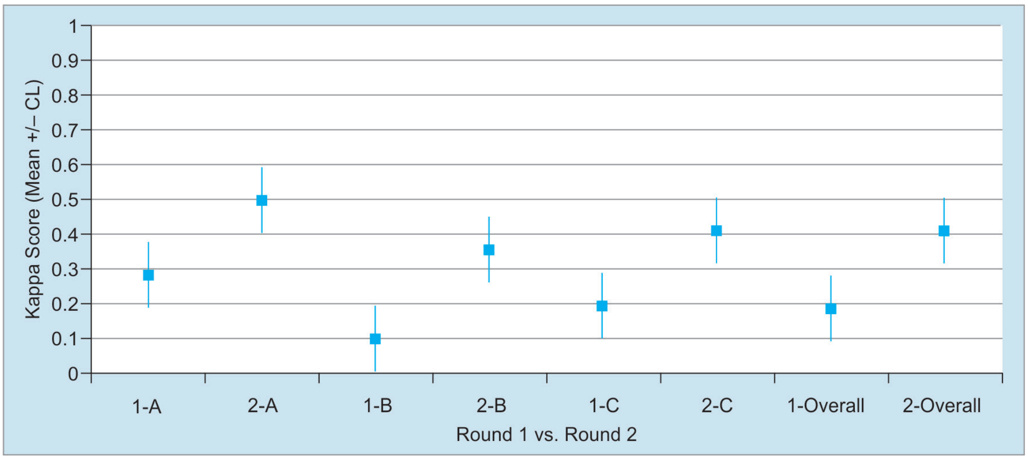
Group II obtained the highest interobserver agreement for type A (0.28) and the lowest for type B (0.10) in round 1. This subgroup obtained the highest interobserver agreement for type A (0.50) and the lowest



Graph 1: Inter-rater reliability



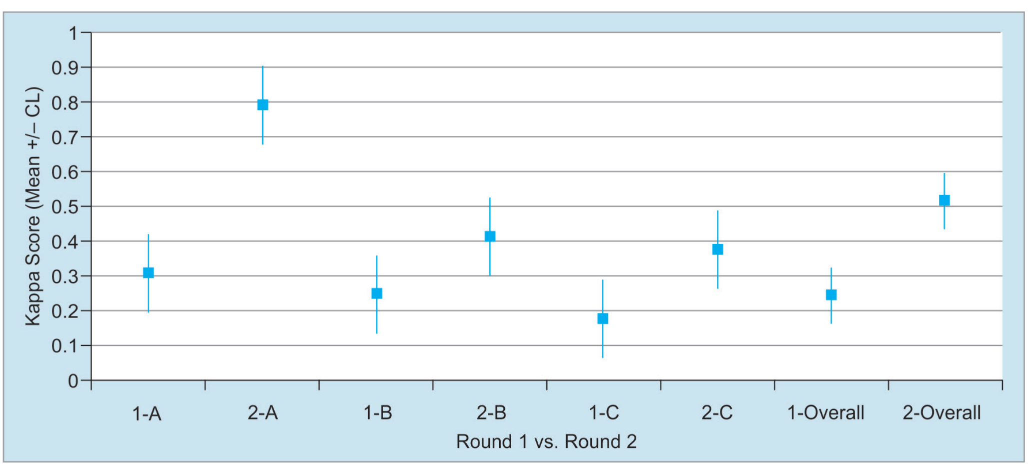
Graph 2: Intratester reliability (attending physicians and fellow)



Graph 3: Intratester reliability (senior residents)

for type B (0.35) in round 2. Comparing rounds 1 to 2, significant improvement was seen in the identification of type A (0.28 *vs* 0.50), type B (0.10 *vs* 0.35), and type C (0.19 *vs* 0.41) (Graph 3). Overall, group II kappa scores for the identification of types significantly improved from rounds 1 to 2 (0.18 *vs* 0.41).

Group III obtained the highest interobserver agreement for type A (0.31) and the lowest for type C (0.17) in round 1. This subgroup obtained the highest interobserver agreement for type A (0.79) and the lowest for type C (0.38) in round 2. Comparing rounds 1 to 2, significant improvement was seen in the identification of type A (0.31 *vs* 0.79), but not of types B (0.25 *vs* 0.41) and C (0.17 *vs* 0.38) (Graph 4). Overall, group III kappa scores for the identification of types significantly improved from rounds 1 to 2 (0.24 *vs* 0.52).



Graph 4: Intratester reliability (junior residents)

DISCUSSION

With the rise in cementless fixation in the United States over the past 20 years, more attention has been devoted to the study of proximal femoral morphology. Previous designs of cementless femoral prosthesis were based on the assumption that age and sex did not affect the shape of the proximal femur.[3] Multiple studies have demonstrated that the geometry of the proximal femur differs significantly with these variables.[3,5-7] As such, implant design has changed to account for this, which should theoretically lead to increased implant longevity and decreased perioperative complications.

With the improvement of implant design and introduction of alternative bearing surfaces, we have seen an increase in younger patients receiving THA. These patients typically will present with "type A" proximal femoral morphology. The increased density noted in this population's proximal femur creates a concern with traditional proximally porous-coated uncemented femoral components. Traditional stems may get "hung up" in the diaphysis which could lead to thigh pain[8] and potentially a lack of osteointegration.[5] Furthermore, the potential for periprosthetic fracture may potentially be increased in this patient population.

In contrast, the aging population tends to have a decrease in the density of the diaphyseal cortices which is commonly referred to as "type C" proximal femoral morphology. These patients seem to be at risk for a poor in-growth potential based on geometrical limitations of their morphology and implant design. It is thought by some that this type of morphology is a relative contraindication to cementless femoral components.

To our knowledge, this is the first study to assess the reliability of classifying proximal femoral morphology as described by Dorr et al.[1] In both rounds of evaluation, type A morphology had the highest interobserver agreement. However, our results show that interobserver agreement was only fair and moderate in rounds 1 and 2. The interobserver agreement for types B and C was slight to fair in both rounds of evaluation, suggesting increased difficulty in the identification of these specific subtypes.

Experience may play a role in the identification and reliability of the specific morphology types. Our results revealed both junior and senior residents showed improvements in specific subtypes between the two rounds of testing. However, the fellowship trained attending physicians and the fellow did not show improvements between rounds in our study.

Based on our findings, we believe the diversity of the proximal femoral morphology likely represents a continuum rather than three distinct shapes. This may imply that for many proximal femoral canals, characterization of a single type may not be possible. Our results suggest this is more likely with patients who fall into the types B and C categories. The clinical significance of this continues to be a challenge in our practice and we treat each individual patient on a case-by-case basis regarding implant selection and determining whether to use cement or press-fit our implants. We concede that we did not evaluate fixation method, implant design, and/or outcomes based on classification in this study. Further research in this area is warranted.

Our study is not without limitations. We had difficulty identifying specific types of proximal femoral morphology that were within the standard measurements as previously described in both the anterior–posterior and lateral views.[1] We believe this likely demonstrates that the femoral canal does not have three distinct shapes but rather presents a continuum as previously discussed. To help eliminate potential error, we had the CI measured by two independent investigators who were not involved in the reliability portion of the study. These measurements needed to be within the reference range based on the original classification in order to be considered for analysis. In addition, although at our institution our radiographs are standardized, slight variability in hip rotation could make a difference in our measurements. The reliability of this study was based on the visual perception of the Dorr classification by the participants and not the radiographic measurements which had previously been done in the original study. Although a potential limitation, we feel this represents real-world application of this classification system. Lastly, all participants were asked to review the original article by Dorr et al defining the classification system of proximal femoral morphology; we did not have a way to verify this portion of the study. Notwithstanding, the brief presentation with examples of each type of proximal femoral morphology given prior to each round of evaluation was done in an attempt to ameliorate this possibility.

Radiographic classification systems are routinely used in orthopedics, but they are seldom studied for their reliability prior to being used in clinical practice. Our results raise the possibility that the morphology of the proximal femur is more challenging to identify than previously described, subject to each surgeon's own level of clinical experience and radiographic interpretation.

ACKNOWLEDGMENT

The authors acknowledge the assistance of Cynthia Green for her assistance in the preparation of this manuscript.

REFERENCES

1. Dorr LD, Faugere MC, Mackel AM, Gruen TA, Bognar B, Malluche HH. Structural and cellular assessment of bone quality of proximal femur. Bone 1993 May-Jun;14(3):231-242.

2. Laine HJ, Lehto MU, Moilanen T. Diversity of proximal femoral medullary canal. J Arthroplasty 2000 Jan;15(1):86-92.
3. Noble PC, Box GG, Kamaric E, Fink MJ, Alexander JW, Tullos HS. The effect of aging on the shape of the proximal femur. Clin Orthop Relat Res 1995 Jul;(316):31-44.
4. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174.
5. Casper DS, Kim GK, Parvizi J, Freeman TA. Morphology of the proximal femur differs widely with age and sex: relevance to design and selection of femoral prostheses. J Orthop Res 2012 Jul;30(7):1162-1166.
6. Husmann O, Rubin PJ, Leyvraz PF, de Roguin B, Argenson JN. Three-dimensional morphology of the proximal femur. J Arthroplasty 1997 Jun;12(4):444-450.
7. Noble PC, Alexander JW, Lindahl LJ, Yew DT, Granberry WM, Tullos HS. The anatomic basis of femoral component design. Clin Orthop Relat Res 1988 Oct;(235):148-165.
8. Kinov P, Radl R, Zacherl M, Leithner A, Windhager R. Correlation between thigh pain and radiological findings with a proximally porous-coated stem. Acta Orthop Belg 2007 Oct;73(5):618-624.