Shafa Ortho J. 2018 February; 5(1):e64801. doi: 10.5812/soj.64801.

Published online 2018 February 17. **Research Article**

Inter and Intra-Observer Reliability of Dorr Classification in Proximal Femur Morphology

Farid Najd Mazhar,[1] Davod Jafari,[1] Marzieh Nojoomi,[2] Alireza Mirzaei,[1,*] and Hamed Tayebi[1]

[1]Bone and Joint Reconstruction Research Center, Shafa Orthopedic Hospital, Iran University of Medical Sciences, Tehran, IR Iran
[2]Preventive Medicine and Public Health Research Center, Department of Community Medicine, School of Medicine, Iran University of Medical Sciences, Tehran, IR Iran

[*]*Corresponding author*: Alireza Mirzaei, Shafa Orthopedic Hospital, Baharestan Sq, Tehran, IR Iran. Tel: +98-2133542000-8, Fax: +98-2133542020, E-mail: mirzaeialireza26@gmail.com

Received 2017 December 09; **Revised** 2018 January 22; **Accepted** 2018 January 23.

Abstract

Background: Proximal femur morphology determines the choice of cemented or uncemented Total Hip Arthroplasty (THR), and Dorr classification is widely used to type this characteristic. As a result, a considerable reliability for such typing system is expected.
Objectives: The current study aimed at assessing inter and intra-observer reliability of Dorr classification in proximal femur morphology.
Methods: A total of 50 standard hip radiographs from patients with trauma or THR candidates were evaluated by 6 observers, including 2 senior orthopedic residents, 2 orthopedic surgeons, and 2 hip fellowship-trained orthopedic surgeons, using the Dorr classification system. Subsequently, radiographs were re-numbered and re-evaluated, 6 weeks after the first review. Cohen's Kappa statistic was used for statistical analysis of inter and intra-observer reliability.
Results: The mean intra-observer and inter-observer kappa value for the residents was 0.560 and 0.592, respectively. The mean intra-observer and inter-observer kappa value for the orthopedists was 0.566 and 0.540, respectively. The mean intra-observer and inter-observer kappa value for the hip fellowships was 0.477 and 0.490, respectively.
Conclusions: A minimum to moderate inter and intra-observer reliability of Dorr classification system was seen between the observers of the current study. These results suggest that the classification of the femur morphology is more challenging than previously described and characterization of a single Dorr type, especially in case of uncertainty between type B and C of Dorr might need reconsideration, in order to prevent its adverse effect on the patients' outcome.

Keywords: Total Hip Arthroplasty, Proximal Femur Morphology, Dorr Classification, Reliability

1. Background

Classification systems aid surgeons to characterize a problem, suggest a potential prognosis, and offer guidance in determining the optimal therapeutic method for a particular condition. In addition, classification systems play a key role in the reporting of clinical and epidemiologic data, allowing uniform comparison and documentation of such conditions.

An optimal classification system should be reliable and valid, while an unreliable classification may result in unnecessary harm to patients (1). As a result, evaluation of intra-observer and inter-observer reliability of classification systems is of critical importance.

As population age increases, a considerable number of patients with osteoporosis will need total hip replacement (THR). Several factors determine the risk of femoral periprosthetic fracture, including the choice of cemented or uncemented stem. Proximal femoral morphology is one of the most important components, which determines the use of cemented or uncemented stem (2). In order to unify the decision-making process, Dorr classification of femoral morphology is widely used (3). In this regard, application of cemented THR for Dorr type C and uncemented THR for Dorr type A and B has been acknowledged (4). Given the importance of Dorr type in the selection of femoral stem (cemented vs. uncemented) and the best type of cementless stem, a considerable reliability for such typing system is expected. To the best of the author's knowledge, to date only one published study has evaluated the reliability of the Dorr classification (5).

2. Objectives

This manuscript aimed at investigating the inter-observer and intra-observer reliability of Dorr typing system in THR candidate with osteoporotic proximal femur.

3. Methods

This study was approved by the institutional review board of the university, where the current study was conducted. A total of 50 standard anteroposterior hip radiographs, including proximal half of the femur from patients with trauma or THR candidates, were used in this study. All patients were referred to the hospital during 2010 to 2014, in order to obtain the required treatment.

The Dorr system was employed to classify proximal femoral morphology. It is an X-ray classification system and consists of 3 distinct patterns of shape and structure of the proximal femur, labeled as A, B, and C. Type A corresponds to a small metaphysis, thick cortex, and high narrowed isthmus. Type B corresponds to a wider metaphysis, thinner cortex, and a tapering but wider isthmus. Finally, Type C corresponds to a wide metaphysis and thin cortex with loss of isthmus constriction (3).

The radiographs were reviewed by 6 observers, including 2 senior orthopedic residents, 2 orthopedic surgeons, and 2 hip fellowship-trained orthopedic surgeons. A schematic description of the Dorr classification was available before and during the reading sessions for observers. Names and identifying marks were covered on the radiographs and they were randomly numbered afterwards. All radiographs were reviewed by each observer. Subsequently, radiographs were re-numbered in order to be re-evaluated, 6 weeks after the first review. The observers were blinded to their own and other's reports.

Kappa value distinguishes true agreement of different observations from agreement due to chance alone. As a result, Cohen's Kappa statistic has been used for statistical analysis of interrater reliability. Accordingly, the results of each observer has been compared to its own second report (intra-observer) and with the results of the other studies (inter-observer), as well (6).

According to this analysis, a range of Kappa values from 0 to 1 was obtained; 0 was regarded as the least and 1 as the highest agreement (Table 1) (6).

Table 1. Interpretation of Cohen's Kappa

Kappa Value	Level of agreement
0 - 0.2	None
0.21 - 0.39	Minimal
0.4 - 0.59	Weak
0.6 - 0.79	Moderate
0.8 - 0.90	Strong
0.9 - 1	Perfect

The IBM SPSS Statistics 22 was used for all statistical analysis and P values of <0.05 were considered statistically significant.

4. Results

The degree of agreement was assessed for all possible pairs, presented as a kappa value. The intra-observer kappa value for resident 1 and 2 was 0.657 ($P < 0.001$) and 0.464 ($P < 0.001$), respectively. The inter-observer kappa value between resident 1 and 2 was 0.496 ($P < 0.001$) and 0.688 ($P < 0.001$) for the first and second observation, respectively.

The intra-observer kappa value for the orthopedic surgeon 1 and 2 was 0.465 ($P < 0.001$) and 0.667 ($P < 0.001$), respectively. The inter-observer kappa value between orthopedic surgeon 1 and 2 was 0.385 ($P < 0.001$) and 0.696 ($P < 0.001$) for the first and second observation, respectively.

The intra-observer kappa value for the orthopedic surgeon with hip fellowship 1 and 2 was 0.635 ($P < 0.001$) and 0.320 ($P = 0.001$), respectively. The inter-observer kappa value between the orthopedic surgeons with hip fellowship 1 and 2 was 0.727 ($P < 0.001$) and 0.254 ($P < 0.004$) for the first and second observation, respectively.

The detailed kappa value for inter-observer and intra-observer reports in addition to the corresponding Confidence Intervals (CI) are demonstrated in Table 2.

Table 2. The Detailed Kappa Value for Inter-Observer and Intra-Observer Reports[a]

Group	Kappa Value	P Value	95% CI
Resident 11 & 12	0.657	< 0.001	(0.566, 0.749)
Resident 21 & 22	0.464	< 0.001	(0.366, 0.562)
Resident 11 & 21	0.496	< 0.001	(0.400, 0.592)
Resident 12 & 22	0.688	< 0.001	(0.600, 0.776)
Orthopedist 11 & 12	0.465	< 0.001	(0.367, 0.563)
Orthopedist 21 & 22	0.667	< 0.001	(0.579, 0.755)
Orthopedist 11 & 21	0.385	< 0.001	(0.287, 0.483)
Orthopedist 12 & 22	0.696	< 0.001	(0.601, 0.791)
Hip fellowship11 & 12	0.635	< 0.001	(0.540, 0.730)
Hip fellowship 21 & 22	0.320	0.001	(0.212, 0.428)
Hip fellowship 11 & 21	0.727	< 0.001	(0.638, 0.816)
Hip fellowship 12 & 22	0.254	< 0.004	(0.150, 0.358)

[a]First number refers to the observer and the second one to the session of the assessment. For example resident 21 means: the first observation of resident number 2.

The mean intra-observer and inter-observer kappa value of the residents was 0.560 and 0.592, respectively. The mean intra-observer and inter-observer kappa value of the orthopedists was 0.566 and 0.540, respectively. The mean intra-observer and inter-observer kappa value of the hip fellowships was 0.477 and 0.490, respectively.

The mean total kappa value (intra-observer + inter-observer) for residents, orthopedists, and hip fellowships was 0.576, 0.553, and 0.484, respectively.

5. Discussion

Classification systems have numerous implications. They should facilitate communication among physicians and assist documentation and research. They should have prognostic value for patients and assist physicians in planning their management (1). Indeed, in the context of treatment, a poorly validated classification will be a biased predictor of patients' outcome, and its application may result in unnecessary harm to patients. In the course of scientific clinical research, poorly validated classifications will result in misclassification of patients and bias the study, making the comparison of patient populations between studies difficult (1). Therefore, it is important to judge whether a classification process "measures what we want it to measure, and how well it does so" before it is widely accepted in clinical practice (1, 7).

Some classifications of musculoskeletal conditions have not proved to be reliable when critically evaluated (8). For example, several studies demonstrated questionable inter-observer reliability of Neer classification system for proximal humerus fractures (9, 10). On the other hand, reliability of some other classification methods, such as Fernandez classification system for distal radius fractures, has been evaluated and approved (11).

As the population ages, the number of hip osteoporotic fractures increase (12-14). Recent studies have demonstrated better outcomes with THR than fracture fixation or partial hip replacement in geriatrics (15, 16). This information has led to increased number of arthroplasty used for the treatment of hip fractures. However, the surgeon needs to understand and address the potential pitfalls that osteoporosis could bring to performing THR. The shape of the osteoporotic femur is different from non-osteoporotic femur by having a wider and thinner cortex (17). This characteristic determines the choice of cemented or uncemented THR, and the Dorr classification is widely used to type this characteristic (3). In this regard, those with type A or B of Dorr are advised to receive uncemented THR, while type C of Dorr is better managed by cemented THR (4). As a result, a considerable inter or intra-observer reliability is expected for Dorr reports to assure unbiased prediction of patients' outcome. It is of note that the surgeons, who use cemented stem for all patients regardless of morphology of their proximal femur do not use the Dorr classification.

To the best of the author's knowledge, there is only one available published article, which has recently evaluated the reliability of visual perception of Dorr classification (5). According to their report, the inter-observer reliability ranged from slight to moderate across testing. They also considered experience as a key player in such classification (5). They concluded that the diversity of proximal femur morphology likely represents a continuum rather than 3 distinct classes, and characterization of a single type may not be possible.

The current results also showed a minimum to moderate inter and intra-observer agreement between the observers regarding the level of qualification. In this regard, the mean intra-observer agreement level was 0.560, 0.566, and 0.477 for residents, orthopedists, and hip fellowships, respectively. In addition, the mean inter-observer agreement level was 0.592, 0.540, and 0.490 for residents, orthopedists, and hip fellowships, respectively.

Surprisingly, the level of inter or intra-observer agreement decreased by higher qualification of observers, showing the lowest agreement level among hip fellowships and higher agreement level among residents. These results were inconsistent with the results of the study of Jenning et al., which reported a higher intra-tester reliability among fellowship trained attending physician compared to junior and senior residents (5). This inconsistency needs to be further addressed in future investigations.

According to the current results, classification of the femur morphology is more challenging than previously described and strongly subjected to the surgeon's level of clinical experience and radiographic interpretation. This could imply that characterization of a single type femoral morphology may not be applicable in all cases. This is more likely in patients, who fall in the Dorr type B and C. In other words, when there is uncertainty between Dorr type A and B, it does not affect the choice of therapeutic approach, while uncertainty between type B and C might lead to the subsequent wrong choice of therapeutic strategy (cemented vs. uncemented arthroplasty) and effect of the patients' outcome and quality of life. In the current study, although in the majority of cases a high inter and intra agreement level was seen, in some cases half of the observations were recorded as type B, while the other half were determined as type C (Figure 1). Considering femoral morphology as a continuum, as Jennings et al. implied (5), in such cases the femoral morphology might be located somewhere in the border of type B and C, which makes the classification process troublesome. This is where the Dorr classification mostly fails to adequately classify femoral morphology and needs a resolution in order to prevent misclassification and its subsequent therapeutic burden.

Recently, Abdulkarim et al. reviewed all randomized

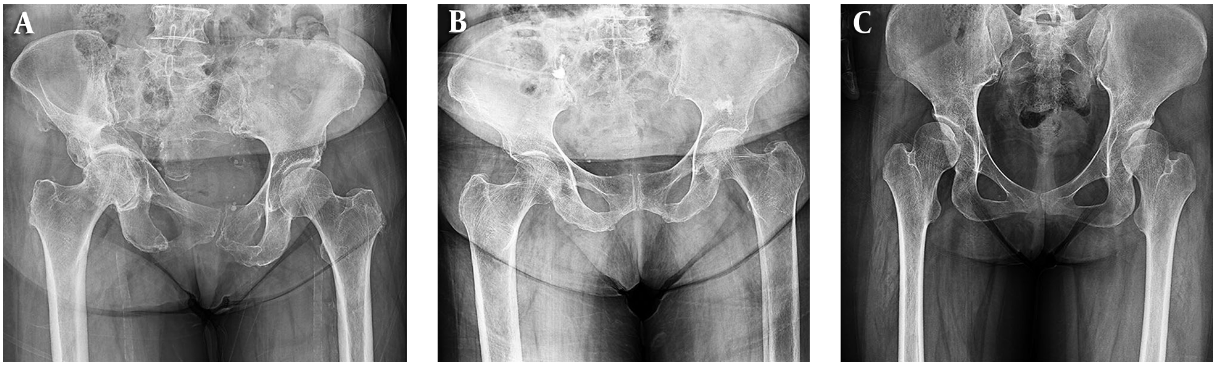

Figure 1. A, A Proximal Femur Radiograph with Complete Agreement on Type B of Dorr Classification; B, A Proximal Femur Radiograph with Complete Agreement on Type C of Dorr Classification; C, A Proximal Femur Radiograph with Half Agreement on Type B and Half Agreement on Type C of Dorr Classification.

controlled trials comparing cemented versus uncemented THRs available in the published literature. According to their report, no significant difference was found between cemented and uncemented THRs in terms of implant survival as measured by the revision rate. However, they concluded that further research, with improved methodology and longer follow up are needed to better define specific subgroups of patients in whom the relative benefits of cemented and uncemented implant fixation can be clearly demonstrated (18). Until then, the Dorr typing system still maintains its importance in the classification of femoral morphology and deciding whether cemented or uncemented stems should be applied.

The primary limitation of the current study was that the number of observers for each category of experience was limited to 2. This number is appropriate using Cohen's Kappa statistic, however, additional observers may be used when the Fleiss Kappa statistic is used. The concern here is that the use of only 2 observers for each category may potentially cause sample error regarding the orthopedic surgeon population. This concern could be mitigated with inclusion of additional observers in future investigations.

In conclusion, more reliable characterization systems are needed to classify femoral morphology and its subsequent choice of therapeutic strategy. In spite of widespread application of Dorr classification system, the current analysis did not confirm the reliability of this system and more evaluation of this system is needed to be performed in future investigations.

References

1. Audige L, Bhandari M, Hanson B, Kellam J. A concept for the validation of fracture classifications. *J Orthop Trauma*. 2005;**19**(6):401–6. [PubMed: 16003200].
2. Mears SC. Management of Severe Osteoporosis in Primary Total Hip Arthroplasty. *Curr Translat Geriatrics Exp Gerontol Rep*. 2013;**2**(2):99–104. doi: 10.1007/s13670-013-0044-7.
3. Dorr LD, Faugere MC, Mackel AM, Gruen TA, Bognar B, Malluche HH. Structural and cellular assessment of bone quality of proximal femur. *Bone*. 1993;**14**(3):231–42. doi: 10.1016/8756-3282(93)90146-2.
4. Marya SKS, Bawari RK. *Total Hip Replacement Surgery: Principles and Techniques*. JP Medical Ltd; 2010.
5. Jennings JM, Thiele RAR, Krynetskiy E, Wellman SS, Attarian DE, Bolognesi MP. ndependent Analysis of the Dorr Classification of. Proximal Femoral Morphology: A Reliability Study. *Duke Orthop J*. 2016;**6**(1):12–6.
6. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;**22**(3):276–82. [PubMed: 23092060].
7. Ackermann C, Lam Q, Linder P, Kull C. Problems in classification of fractures of the proximal humerus. *Zeitschrift fur Unfallchirurgie, Versicherungsmedizin und Berufskrankheiten*. 1986;**79**(4):209.
8. Garbuz DS, Masri BA, Esdaile J, Duncan CP. Classification systems in orthopaedics. *J Am Acad Orthop Surg*. 2002;**10**(4):290–7. [PubMed: 15089078].
9. Brien H, Noftall F, MacMaster S, Cummings T, Landells C, Rockwood P. Neer's Classification System. *J Trauma Injury Infect Crit Care*. 1995;**38**(2):257–60. doi: 10.1097/00005373-199502000-00022.
10. Sidor ML, Zuckerman JD, Lyon T, Koval K, Cuomo F, Schoenberg N. The Neer classification system for proximal humeral fractures. An assessment of interobserver reliability and intraobserver reproducibility. *J Bone Joint Surg*. 1993;**75**(12):1745–50. doi: 10.2106/00004623-199312000-00002.
11. Naqvi SG, Reynolds T, Kitsis C. Interobserver reliability and intraobserver reproducibility of the Fernandez classification for distal radius fractures. *J Hand Surg Eur Vol*. 2009;**34**(4):483–5. doi: 10.1177/1753193408101667. [PubMed: 19321533].
12. Cheng SY, Levy AR, Lefaivre KA, Guy P, Kuramoto L, Sobolev B. Geographic trends in incidence of hip fractures: a comprehensive literature review. *Osteoporosis Int*. 2011;**22**(10):2575–86. doi: 10.1007/s00198-011-1596-z.
13. Cooper C, Cole ZA, Holroyd CR, Earl SC, Harvey NC, Dennison EM, et al. Secular trends in the incidence of hip and other osteoporotic fractures. *Osteoporosis Int*. 2011;**22**(5):1277–88. doi: 10.1007/s00198-011-1601-6.
14. Kanis JA, Odén A, McCloskey EV, Johansson H, Wahl DA, Cooper C. A systematic review of hip fracture incidence and probability of fracture worldwide. *Osteoporosis Int*. 2012;**23**(9):2239–56. doi: 10.1007/s00198-012-1964-3.
15. Blomfeldt R. Comparison of Internal Fixation with Total Hip Replacement for Displaced Femoral Neck Fracturesaid="1022161", Controlled Trial Performed at Four Years</sbt>. *J Bone Joint Surg*. 2005;**87**(8):1680. doi: 10.2106/jbjs.d.02655.

16. Hedbeck CJ, Enocson A, Lapidus G, Blomfeldt R, Törnkvist H, Ponzer S, et al. Comparison of Bipolar Hemiarthroplasty with Total Hip Arthroplasty for Displaced Femoral Neck Fractures. *J Bone Joint Surg*. 2011;**93**(5):445–50. doi: 10.2106/jbjs.j.00474.
17. Ahlborg H, Johnell O, Karlsson M. An age-related medullary expansion can have implications for the long-term fixation of hip prostheses. *Acta Orthopaedica Scandinavica*. 2009;**75**(2):154–9. doi: 10.1080/00016470412331294405.
18. Abdulkarim A, Ellanti P, Motterlini N, Fahey T, O'Byrne JM. Cemented versus uncemented fixation in total hip replacement: a systematic review and meta-analysis of randomized controlled trials. *Orthoped Rev*. 2013;**5**(1):8. doi: 10.4081/or.2013.e8.