



## Original article

# Reproducibility of the Dorr classification and its quantitative indices on plain radiographs

Ryota Nakaya<sup>a</sup>, Masaki Takao<sup>a,\*</sup>, Hidetoshi Hamada<sup>b</sup>, Takashi Sakai<sup>b</sup>, Nobuhiko Sugano<sup>a</sup>

<sup>a</sup> Departments of Orthopaedic Medical Engineering, Osaka University Graduate School of Medicine, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan

<sup>b</sup> Departments of Orthopaedic Surgery, Osaka University Graduate School of Medicine, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan



## ARTICLE INFO

## Article history:

Received 17 November 2017

Accepted 14 November 2018

## Keywords:

Dorr classification  
Reproducibility  
Cortical index  
Canal-to-calcar ratio  
Canal flare index

## ABSTRACT

**Background:** The Dorr classification is widely used to evaluate femoral bone quality, but it has no clear quantitative criteria. This study aimed to evaluate the reproducibility of the Dorr classification and examine its quantitative indices on plain radiographs, which are suitable for objective classification.

**Hypothesis:** Reproducibility of the Dorr classification is influenced by the clinical experience of the examiners, and radiographic indices are required for this classification.

**Materials and methods:** One hundred and one patients were examined using their preoperative plain antero-posterior and lateral radiographs. To evaluate the reproducibility of the Dorr classification, the Dorr type of each patient was judged twice each by three expert hip surgeons and three junior hip surgeons. Indices measured using the plain radiographs were canal-to-calcar ratio, cortical index (CI), and canal flare index. A receiver operating characteristic curve was used to evaluate which measured parameters were suitable as indices for the Dorr classification which was determined by the consultation among three expert hip surgeons.

**Results:** Regarding intra-examiner reproducibility, kappa coefficients for the three junior hip surgeons were 0.36, 0.62, and 0.65, whereas those for the three expert hip surgeons were 0.70, 0.86, and 0.87. Regarding inter-examiner reproducibility, the kappa coefficient for the junior hip surgeons was 0.32, whereas that for the expert hip surgeons was 0.52. The CI on the lateral radiograph had the largest area under the curve (AUC) between types A and B, whereas the CI on the anteroposterior radiograph had the largest AUC between types B and C. The respective cutoff points of the CI on the anteroposterior radiograph were 0.58 between types A and B and 0.49 between types B and C. The respective cutoff points of CI on the lateral radiograph were 0.45 between types A and B and 0.28 between types B and C.

**Conclusion:** The intra-examiner reproducibility of the Dorr classification ranged from “fair” to “almost perfect”, whereas the inter-examiner reproducibility ranged from “fair” to “moderate”. Both were influenced by the level of clinical experience of the examiners. The most suitable index for classification using plain radiographs of the hip is the CI on anteroposterior and lateral radiographs.

**Level of evidence:** IV, retrospective study.

© 2018 Elsevier Masson SAS. All rights reserved.

## 1. Introduction

Femoral bone quality is an important factor that affects periprosthetic fractures and femoral component fixation in total hip arthroplasty (THA) [1–5]. Therefore, preoperative evaluation of femoral bone quality is important when choosing a design and fixation method of femoral component. The Dorr classification is widely used to evaluate femoral bone quality [1]. In this classification, femoral bone quality is divided into three types using anteropos-

terior and lateral plain radiographs of the hip joint. Type A is a sufficiently thick bone cortex in anteroposterior and lateral views. Type C is an extremely thin bone cortex with a stovepipe shape in anteroposterior view and a widened medullary cavity and indistinct cortical bone in lateral view. Type B is between types A and C and comprises a thin cortical bone with a residual funnel shape in anteroposterior view and an indistinct bone cortex at the posterior of the femur with a nonuniform line on the medullary cavity side of the bone in lateral view (Fig. 1). However, the Dorr classification merely classifies the appearance on plain radiographs and has no clear quantitative criteria. Currently, there are only few studies investigating intra-examiner and inter-examiner reproducibility of the Dorr classification [1,6].

\* Corresponding author.

E-mail address: [masaki-tko@umin.ac.jp](mailto:masaki-tko@umin.ac.jp) (M. Takao).

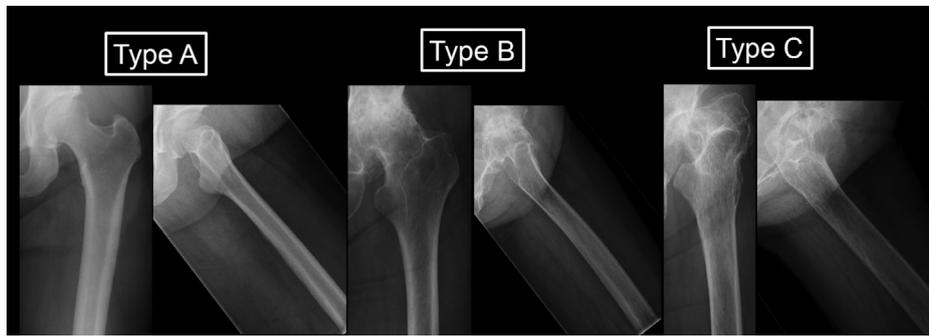


Fig. 1. Plain anteroposterior and lateral radiographs of each Dorr type are shown. The cortex contour is indistinct in the lateral view of the type B and C femora.

We hypothesized that the reproducibility of the Dorr classification is influenced by the clinical experience of the examiners, and radiographic indices are required for this classification. This study aimed to evaluate the reproducibility of the Dorr classification and examine its quantitative indices on plain radiographs, which are suitable for objective classification.

## 2. Patients and methods

### 2.1. Patients

The subjects were 101 consecutive patients who had primary THA at our hospital. Patients with a history of hip fracture surgeries, bone tumors, Grade III or IV developmental dysplastic hips according to the Crowe Classification [7], and indistinct plain radiographs were excluded. There were 78 women and 23 men aged 29–87 years (mean, 60 years) with underlying diseases of osteoarthritis (77), osteonecrosis (16), rapidly destructive coxarthrosis (5), and rheumatoid arthritis (3). This study was approved by the institutional review board of our institution (clinical image analysis of patients with joint disorders, No. 15056-2).

### 2.2. Methods

To evaluate the reproducibility of the Dorr classification, the Dorr type of each patient was judged twice each by three expert hip surgeons and three junior hip surgeons, with a 1-month interval between judgements. Each examiner made their decision after receiving a detailed explanation regarding the Dorr type

characteristics using plain radiographs. The three expert hip surgeons determined the gold standard for the Dorr classification. Indices measured using the plain radiographs were canal-to-calcar ratio (CC ratio) [1], cortical index (CI) [1], and canal flare index (CFI) (Fig. 2) [8]. The CC ratio and CI (anteroposterior and lateral views) were measured using the method described by Dorr et al. [1], and CFI was measured using the method described by Noble et al. [8]. CC ratio: Measured on anteroposterior plain radiographs of the hip. Two points are established on the medial and lateral aspects of the medullary cavity 10 cm distal from the center of the lesser trochanter, and another two points are established on the medial and lateral aspects of the medullary cavity 3 cm distal from the center of the lesser trochanter. The CC ratio was obtained by dividing the distance  $x$  of the medullary cavity at the point 10 cm distal from the center of the lesser trochanter by the distance  $y$  at the point at the center of the lesser trochanter where the two lines formed by the two points on the internal aspect and the two points on the lateral aspect of the medullary cavity pass. CI was measured on anteroposterior and lateral plain radiographs of the hip. CI was obtained by dividing the thickness of the femoral bone cortex at the point 10 cm distal from the center of the lesser trochanter (distance of the lateral aspect of the femoral bone cortex  $z$  – distance of the medullary cavity  $x$ ) by the distance of the lateral aspect of the femoral bone cortex  $z$ . In this study, CI measured on an anteroposterior radiograph was abbreviated as CI (AP) and CI measured on a lateral radiograph was abbreviated as CI (lat). CFI: Measured on anteroposterior plain radiographs of the hip. The width of the medullary cavity proximal from the lesser trochanter by 35% of the vertical distance from the lesser trochanter to the femoral head center was designated  $a$ . CFI was obtained by dividing  $a$  by the

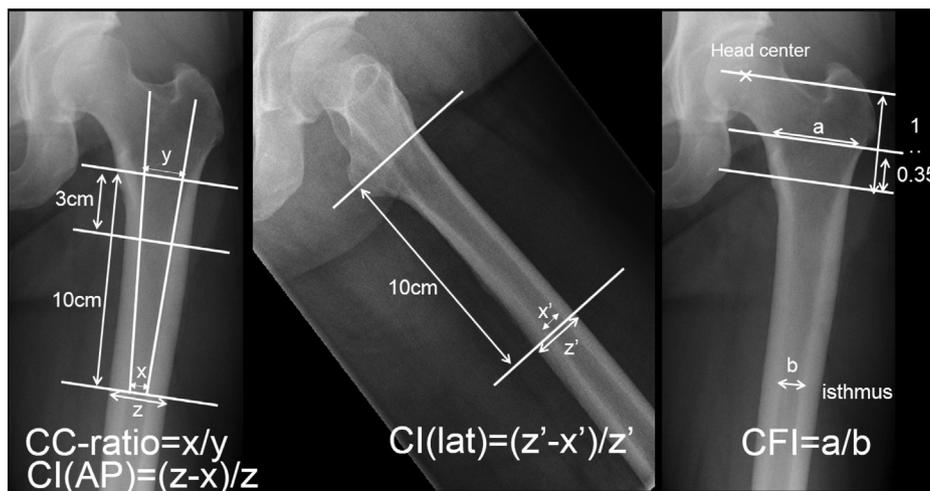


Fig. 2. Methods of measuring each parameter using plain radiographs are shown. CC ratio: canal-to-calcar ratio; CI: cortical index; CFI: canal flare index.

**Table 1**  
Intra-examiner and inter-examiner reproducibility of the Dorr classification.

Examiner	Fleiss' kappa coefficient (95% CI)	Degree of agreement <sup>a</sup>
<b>Intra-examiner reproducibility</b>		
Junior 1	0.65 (0.48–0.81)	Substantial
Junior 2	0.36 (0.18–0.54)	Fair
Junior 3	0.62 (0.47–0.77)	Substantial
Expert 1	0.70 (0.55–0.86)	Substantial
Expert 2	0.87 (0.69–1.06)	Almost perfect
Expert 3	0.86 (0.69–1.02)	Almost perfect
<b>Inter-examiner reproducibility</b>		
Junior surgeons	0.32 (0.22–0.41)	Fair
Expert surgeons	0.52 (0.43–0.62)	Moderate
Overall	0.44 (0.40–0.49)	Moderate

<sup>a</sup> Interpretation of kappa is according to Landis & Koch [10].

width of the isthmus of the medullary cavity *b*. All radiographic measurements were performed by an examiner who was not involved in the determination of the gold standard of the Dorr classification.

### 2.3. Method of assessment and statistical analysis

Statistical analyses were performed using IBM SPSS Statistics 23. The Fleiss' kappa coefficient was used to evaluate the reproducibility of the Dorr classification [9]. The degree of agreement was assessed using the Landis and Koch's interpretation of the kappa coefficient [10].  $\kappa$  coefficient <0 was defined as "poor,"  $\geq 0$  but  $\leq 0.20$  as "slight,"  $\geq 0.21$  but  $\leq 0.40$  as "fair,"  $\geq 0.41$  but  $\leq 0.60$  as "moderate,"  $\geq 0.61$  but  $\leq 0.80$  as "substantial," and  $\geq 0.81$  as "almost perfect."

A receiver operating characteristic (ROC) curve was used to determine the parameters on plain radiographs, which are suitable to be used as indices for the Dorr classification [11,12]. The optimal cutoff point of each radiographic index was calculated using the point on the ROC curve closest to (0, 1). The intraexaminer and interexaminer reproducibility of the Dorr classification, based on the new index on plain radiographs, was assessed by a junior hip

surgeon (junior 1) and an expert hip surgeon (expert 1). The level of significance was set at a *p*-value <0.05.

### 3. Results

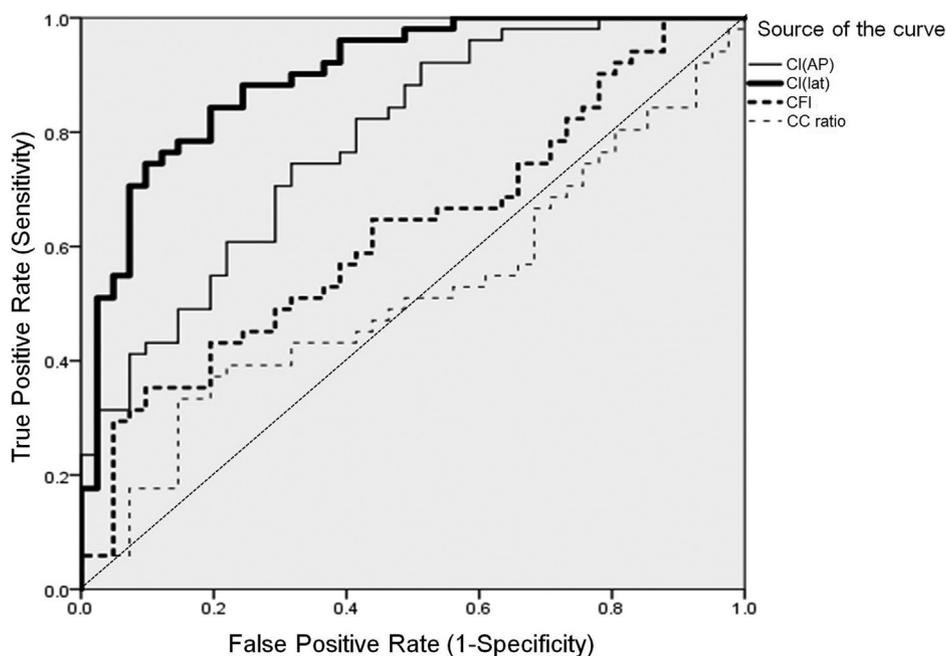
Regarding intra-examiner reproducibility, the kappa coefficients for the three junior hip surgeons were 0.36, 0.62, and 0.65, whereas those for the three expert hip surgeons were 0.70, 0.86, and 0.87. Regarding inter-examiner reproducibility, the kappa coefficients for the junior hip surgeons, expert hip surgeons, and all surgeons were 0.32, 0.52, and 0.44, respectively (Table 1).

The Dorr type gold standard was type A in 51 patients, type B in 41, and type C in nine. The evaluation using ROC curves revealed an area under the curve (AUC) between types A and B of 0.518 for CC ratio, 0.784 for CI (AP), 0.899 for CI (lat), and 0.626 for CFI (Fig. 3 and Table 2). The optimum cutoff value for CI (lat), which had the largest AUC between types A and B, was 0.45. AUC between types B and C was 0.694 for CC ratio, 0.925 for CI (AP), 0.892 for CI (lat), and 0.883 for CFI (Fig. 4 and Table 2). The optimum cutoff value for CI (AP), which had the largest AUC between types B and C, was 0.49.

The intra-examiner and inter-examiner reproducibility of the Dorr classification based on the CI (AP) and CI (lat) is listed in Table 3. The degree of agreement of the intra-examiner reproducibility improved compared with that observed using the conventional classification except for the CI (lat) of the expert hip surgeon (Table 1). Similarly, the degree of agreement of the inter-examiner reproducibility improved when the CI (AP) was used as the reference radiographic index for the classification.

### 4. Discussion

In the current study, the intra-examiner reproducibility of the Dorr classification for junior hip surgeons was "fair" (one surgeon) and "substantial" (two surgeons). For the expert surgeons, the intra-examiner reproducibility was "substantial" (one surgeon) and "almost perfect" (two surgeons). Overall, the inter-examiner reproducibility of the Dorr classification was "fair" in the junior hip surgeons group and "moderate" in the expert hip surgeons group. This indicated that both the intra-examiner and inter-examiner

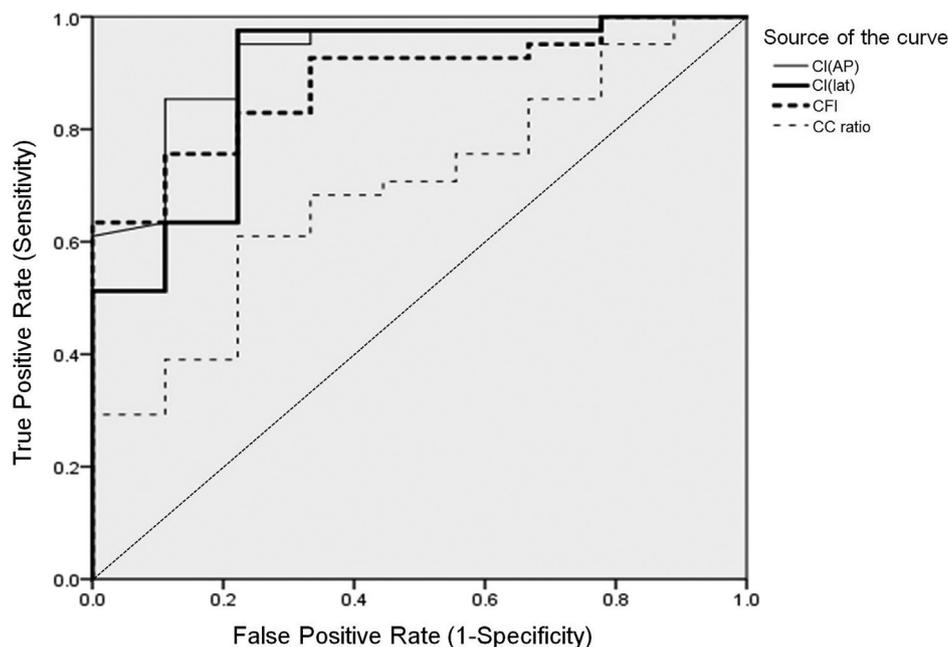


**Fig. 3.** Receiver operating characteristic curves of each parameter on plain radiographs for differentiating between types A and B.

**Table 2**  
Areas under the curve (AUC) and cutoff points of each radiographic parameter.

Radiographic parameters	Types A vs. B			Types B vs. C		
	AUC (95% CI)	p-Value	Cutoff	AUC (95% CI)	p-Value	Cutoff
CC ratio	0.518(0.399–0.637)	$p = 0.76$	0.52	0.694(0.520–0.868)	$p = 0.07$	0.60
CI (AP)	0.784(0.692–0.876)	$p < 0.01$	0.58	0.925(0.837–1.000)	$p < 0.01$	0.49
CI (lat)	0.899(0.836–0.962)	$p < 0.01$	0.45	0.892(0.768–1.000)	$p < 0.01$	0.28
CFI	0.626(0.513–0.740)	$p = 0.04$	4.38	0.883(0.783–0.984)	$p < 0.01$	3.87

CC ratio: canal-to-calcar ratio; CFI: canal flare index; CI: cortical index; AP: anteroposterior view; lat: lateral views; 95% CI: 95% confidence interval.



**Fig. 4.** Receiver operating characteristic curves of each parameter on plain radiographs for differentiating types B and C.

**Table 3**  
Intra-examiner and inter-examiner reproducibility of the Dorr classification according to the cortical index on anteroposterior and lateral radiographs.

Index on plain radiograph	Examiner	Fleiss kappa coefficient(95% CI)	Degree of agreement <sup>a</sup>
Intra-examiner reproducibility			
CI (AP)	Junior 1	0.89 (0.74–1.04)	Almost perfect
	Expert 1	0.86 (0.72–1.00)	Almost perfect
CI (lat)	Junior 1	0.81 (0.65–0.97)	Almost perfect
	Expert 1	0.78 (0.62–0.94)	Substantial
Inter-examiner reproducibility			
CI (AP)	Junior 1 and expert 1	0.68 (0.58–0.83)	Substantial
CI (lat)	Junior 1 and expert 1	0.59 (0.43–0.75)	Moderate

CI: cortical index; AP: anteroposterior view; lat: lateral views; 95% CI: 95% confidence interval.

<sup>a</sup> Interpretation of kappa is according to Landis & Koch [10].

reproducibilities of the Dorr classification are influenced by the level of clinical experience of the examiners. In addition, even for the expert hip surgeons, the inter-examiner reproducibility was “moderate”, thus implying that objective and quantitative indices are required to properly judge the Dorr classification in the clinical setting. Jennings et al. assessed the inter-observer reliability of the Dorr classification using the Fleiss' kappa coefficient among adult reconstruction fellowship trained attending surgeons and fellows as well as among senior residents and junior residents and reported “slight” to “moderate” inter-observer reliability, a finding consistent with the results of the present study [6].

The Dorr classification is a method of qualitative classification that is determined using anteroposterior and lateral plain radiographs of the hip joint. In our exploration of quantitative indices that are suitable for the Dorr classification, we found parameters that were easily measured on plain radiographs. We evaluated

CI, an index of cortical bone thickness of the distal shaft of the femur [1]; CC ratio, an index of the shape of the medullary cavity of the proximal femur [1]; and CFI, an index of the flare shape of the medullary cavity of the proximal femur [8,13]. Evaluation using ROC curves revealed that CI (lat) between types A and B and CI (AP) between types B and C were the indices that could be most sensitively judged. Originally, types A and B were differentiated according to the presence or absence of indistinctness of the femoral bone cortex on lateral plain radiographs of the hip, and the result of this differentiation reflected the definition of each type. Types B and C are differentiated according to the widening of the medullary cavity and thinness of the bone cortex in anteroposterior and lateral views. Our results also showed a larger AUC for CI than that for other parameters in anteroposterior and lateral views. AUC for CI was greater in the anteroposterior than in the lateral view.

To assess the reliability of the CI (AP) and CI (lat) as radiographic indices for the Dorr classification, a junior hip surgeon and an expert hip surgeon measured them twice for all cases. The intra-examiner reproducibility of both hip surgeons improved from “substantial” to “almost perfect,” except for the CI (lat) measured by the expert hip surgeon. Moreover, the inter-examiner reproducibility improved from “moderate” to “substantial” when the CI (AP) was used as the reference; however, this improvement was not observed when the CI (lat) was used. This may be attributed to the lower clarity of the cortex contour on lateral radiographs (Fig. 1). In addition, the inter-examiner reproducibility of the Dorr classification did not reach the level of “almost perfect” agreement, even when the CI (AP) was used. This may be due to variations in the proximal femoral morphology being a continuum, and it is inherently difficult to classify it into three specific types [6].

The present study has three limitations. First, it involves a small sample size. Sim J et al. reported that the number of subjects required in a 2-rater study to detect a statistically significant kappa ( $p \leq 0.05$ ) on a dichotomous variable, where it is 50 with 80% power regardless of the proportions of positive diagnoses, assuming the null hypothesis value of kappa to be zero [14]. Regarding the ROC curve analysis, power was calculated using R version 3.5.0 and was shown to be >80% in both the CI (AP) and CI (lat). Thus, the sample size of the current study (101 cases) is considered sufficient. Second, the effect of the femoral rotation during X-ray imaging on the reproducibility of Dorr classification was not considered. The clarity of the cortex contour may change with the rotation of the femur, and the effect of femoral rotation on the clearness of the cortex contour may differ among the three types. The effect of femoral rotation on the reproducibility of the Dorr classification should be evaluated using three-dimensional computed tomography. Third, the correlation between improvement in the Dorr classification and bone quality using dual-energy X-ray absorptiometry (DEXA) or quantitative computed tomography was not evaluated. Sah et al. [15] reported regarding the degree to which parameters on plain radiographs (Dorr classification, CI, CC ratio, and Singh index) were related to DEXA T-score (femoral neck, greater trochanter of the femur, spinal column) in 32 patients with osteoarthritis. They also found that when osteoporosis had a T-score of  $-2.5$ , the CI (AP) cutoff was 0.50, which was almost the same as our cutoff of 0.49 between types B and C. If we consider that Type C is an index of poor bone quality [1,15], our cutoff could suitably determine Type C as an index of poor bone quality when the index is used for preoperatively evaluating femoral bone quality.

In conclusion, the intra-examiner reproducibility of the Dorr classification ranged from “fair” to “almost perfect” and the inter-examiner reproducibility ranged from “fair” to “moderate”. Both were influenced by the level of clinical experience of the examiners. The most suitable index for classification using plain radiographs of the hip is the CI on anteroposterior and lateral radiographs. The respective cutoff points of the CI on the anteroposterior radiograph were 0.58 between types A and B and 0.49 between types B and

C. The respective cutoff points of CI on the lateral radiograph were 0.45 between types A and B and 0.28 between types B and C.

### Disclosure of interest

The authors declare that they have no competing interest.

### Funding

No financing was received for this study.

### Contribution

All authors had made substantial contributions to all of the following:

- the conception and design of the study, or acquisition of data, or analysis and interpretation of data;
- drafting the article or revising it critically for important intellectual content,
- final approval of the version to be submitted.

### References

- [1] Dorr LD, Faugere MC, Mackel AM, Gruen TA, Bogner B, Malluche HH. Structural and cellular assessment of bone quality of proximal femur. *Bone* 1993;14:231–42.
- [2] Kobayashi S, Saito N, Horiuchi H, Iorio R, Takaoka K. Poor bone quality or hip structure as risk factors affecting survival of total-hip arthroplasty. *Lancet* 2000;355:1499–504.
- [3] Lindahl H. Epidemiology of periprosthetic femur fracture around a total hip arthroplasty. *Injury* 2007;38:651–4.
- [4] Sychterz CJ, Engh CA. The influence of clinical factors on periprosthetic bone remodeling. *Clin Orthop Relat Res* 1996;322:285–92.
- [5] Russell LA. Osteoporosis and orthopedic surgery: effect of bone health on total joint arthroplasty outcome. *Curr Rheumatol Rep* 2013;15:371.
- [6] 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;6:12–6.
- [7] Crowe JF, Mani VJ, Ranawat CS. Total hip replacement in congenital dislocation and dysplasia of the hip. *J Bone Joint Surg Am* 1979;61:15–23.
- [8] 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;316:31–44.
- [9] Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971;76:378–82.
- [10] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- [11] Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med* 2003;29:1043–51.
- [12] Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285–93.
- [13] Noble PC, Alexander JW, Lindahl LJ, Yew DT, Granberry WM, Tullos HS. The anatomic basis of femoral component design. *Clin Orthop Relat Res* 1988;235:148–65.
- [14] Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257–68.
- [15] Sah AP, Thornhill TS, LeBoff MS, Glowacki J. Correlation of plain radiographic indices of the hip with quantitative bone mineral density. *Osteoporos Int* 2007;18:1119–26.