Araştırma Makalesi / *Research Article*

İzmir Eğitim ve Araştırma Hastanesi Tıp Dergisi (Medical Journal of İzmir Hospital) 23 (2): 147-153, 2019

TRAVMATİK AYAK BİLEĞİ OSTEOARTRİTİ İÇİN KULLANILAN SINIFLANDIRMA SİSTEMLERİNİN GÖZLEMCİLER ARASI VE GÖZLEMCİ İÇİ GÜVENİLİRLİK ANALİZİ

ANALYSIS OFINTEROBSERVER AND INTRA-OBSERVER RELIABILITY OF CLASSIFICATION SYSTEMS USED FOR POST-TRAUMATIC ANKLE OSTEOARTHRITIS

Adil TURAN Cemil AKTAN

Sağlık Bilimleri Üniversitesi, Antalya Eğitim ve Araştırma Hastanesi, Ortopedi ve Travmatoloji Kliniği, Antalya

Anahtar Sözcükler: Ayak bileği yaralanmaları, osteoartrit, sınıflandırma, güvenilirlik, gözlemciler arası değişkenlik, gözlemci içi değişkenlik

Keywords: Ankle injury, osteoarthritis, classification, reliability, interobserver variation, intra-observervariation

Yazının alınma tarihi:26.12.2018 Kabul tarihi:03.02.2109 Online basım:14.03.2019

ÖZ

Giriş: *Bu çalışmanın amacı, ayak bileği ekleminin osteoartritinde (OA) üç farklı radyografik sınıflandırma sisteminin gözlemciler arası ve gözlemci içi güvenilirliklerini araştırmaktır. Ayrıca radyografik evre ile hastaların klinik durumu arasındaki korelasyon analiz edildi.*

Gereç ve Yöntem: *Ayak bileği kırığı nedeniyle opere edilen 60 hasta çalışmaya dahil edildi. Tüm hastalar en az 12 ay takip edildi ve klinik olarak AOFAS skoru ile değerlendirildi. İki uzman ortopedist, bu hastaların ayak bileği radyografilerini Kellgren-Lawrence (KL), Takakura ve van Dijk ayak bileği osteoartriti sınıflandırma sistemlerine göre sınıflandırdı. Değerlendirmeler, her gözlemci tarafından en az 4 hafta arayla iki ayrı oturumda rastgele olarak yapıldı. Kappa istatistikleri, iki gözlemci için gözlemciler arasında ve aynı gözlemcininfarklı değerlendirmeleri arasında rölatif bir kabul edilebilirlik düzeyi oluşturmak için kullanıldı.*

Bulgular: *KL sınıflandırması için gözlemci içi güvenilirlik her iki gözlemci için de orta düzeyde bulundu (sırasıyla κ: 0.277 ve κ: 0340). KL için gözlemciler arası güvenilirlik ilk değerlendirmedeorta (κ: 0.362), ikinci değerlendirmedezayıf (κ: 0.106) idi. Takakura sınıflandırmasının gözlemci içi güvenilirliği gözlemci A için orta (κ: 0.255), gözlemci B için makul (κ: 0.576), gözlemciler arası güvenilirliği ise için ilk değerlendirmede orta düzeyde (κ: 0.472) ikinci degerlendirmede ise zayıf (κ: 0.114) olarak bulundu. Van Dijk sınıflandırmasının gözlemci içi güvenilirliği gözlemci A için orta (κ: 0.321) ve gözlemci B için makul (κ: 0.443) gözlemciler arası güvenilirliği ise ilk değerlendirmede orta (κ: 0.328), ikinci değerlendirmede ise zayıf (κ:0.163) olarak bulunduTüm sınıflandırma sistemleri ile AOFAS arasında anlamlı bir ilişki vardı.*

Sonuç: *Çalışmada değerlendirilen ayak bileği OA sınıflandırma sistemlerinden hiçbiri için kabul edilebilir güvenilirlik düzeyleri elde edilmedi (κ> 0.80). Bununla birlikte, bu üç ölçek arasında, Takakura sınıflandırmasının nispeten daha fazla güvenilirlik ve hastaların klinik durumları ile daha iyi korelasyon gösterdiği bulundu.*

SUMMARY

Introduction: *The aim of this study was to investigate the interobserver and intra-observer reliabilities of three different radiographic grading scales of osteoarthritis (OA) of the ankle joint. Furthermore, correlation between the radiographic grade and the clinical status of the patients were analyzed.*

Material and Method: *Sixty patients who underwent operative treatment for ankle fracture were included to this study. All patients were followed at least 12 months and clinically evaluated with AOFAS score. Two consultant orthopedic surgeons classified ankle radiographs of these patients according to the Kellgren-Lawrence (KL), Takakura and van Dijk ankle osteoarthritis grading scale. Assessments were performed in random order by each observer on two separate occasions, at least 4 weeks apart. Kappa statistics were used to establish a relative level of agreement between observers for the two readings and between separate readings by the same observer.*

Results: *Intra-observer reliability for KL was fair for both observers (κ: 0.277, and κ: 0340 respectively). Interobserver reliability for KL was fair (κ:0.362) at first occasion, slight (κ:0.106) at the second occasion. Intra-observer reliability of Takaura classification was fair (κ: 0.255) for observer A, and moderate (κ: 0.576) for observer B. Interobserver reliability for Takakura classification was moderate (κ: 0.472) at first occasion, slight (κ: 0.114) at the second occasion. Intra-observer reliability of van Dijk classification was fair (κ:0.321) for observer A, and moderate (κ:0.443) for observer B. Interobserver reliability for van Dijk classification was fair (κ:0.328) at first occasion, slight (κ:0.163) at the second occasion. There was a significant correlation between all classification schemes and AOFAS.*

Conclusions: *None of the studied ankle OA grading scales showed acceptable reliability (κ > 0.80). However, among these three scales, Takakura classification showed relatively greater reliability, and better correlation with the clinical status of the patients.*

INTRODUCTION

Osteoarthritis (OA) of the ankle joint is a chronic disease which is characterized by degenerative changes in articular cartilage, degradation of periarticular bone and surrounding soft tissues. Although ankle osteoarthritis (OA) is less frequent than knee and hip joint OA, approximately 1% of the world's adult population is affected by this debilitating disorder. Several factor may play role in the etiology of ankle OA, however traumatic ankle injuries, particularly ankle fractures and dislocations, are the leading cause of ankle OA (1).

Management of ankle OA comprise various treatment methods ranging from conservative measures to advanced surgical options such as total ankle arthroplasty. Choosing appropriate treatment method for a particular patient depends both on the clinical and radiographic evaluations (2,3). Radiographs are crucial to determine the severity of OA, monitoring the progression of the disease, and assessment of the outcomes. Currently, several radiographic grading scales of OA of the ankle joint have been proposed, such as Kellgren–Lawrence (KL), Takakura and Van Dijk classifications. Almost all of these classification systems are primarily based on the subjective ratings of the observers about the narrowing of the joint space, and secondary findings of OA such as osteophytes (4,5,6). Any classification system used in clinical practice should be reliable and reproducible. Although these classification systems are widely used, studies about the reliability of these OA grading scales are limited in the current literature (7,8). The aim of this study was to investigate the interobserver and intra-observer reliabilities of three different radiographic grading scales of OA of the ankle joint. Furthermore, correlation between the radiographic grade and the clinical status of the patients were analyzed.

MATERIAL AND METHOD

Patients

A retrospective review was performed on 60 patients who underwent open reduction and internal fixation for ankle fracture between 2014 and 2016 in our institution. There were 38 male and 22 female patients with a mean age of 47±15.8 (range, 18-81). All patients were followed up at least 12 months, with a mean of 23.7±8.2 months (range, 12-39). There were 9

(%15) pronation-adduction (PAD) type fracture, 9 (15%) pronation-external rotation (PER), 9 (15%) supination-adduction (SAD), and the remaining 33 (%55) had supination-external rotation (SER) type fracture. At the final follow-up, functional outcome was assessed with The American Orthopedic Foot and Ankle Score (AOFAS). Radiological evaluation was performed with anteroposterior and lateral ankle radiographs. This study was carried out in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments (as revised in Brazil in 2013).

Inter and intra-observer assessments

Antero-posterior and lateral ankle radiographs of the patients at the final follow-up were obtained from the picture archiving and communication system (PACS). All identifying marks were removed from the films, which were then sorted and labeled. Two consultant orthopedic surgeons who were familiar with the ankle OA classifications took part in the study. Before initiation of the study, to provide an agreement on criteria for radiographic evaluation, each observer was provided with the same descriptions and diagrams of the classification systems (Table 1, 2 and 3). Radiologic assessments were performed in random order by each observer on two separate occasions, at least 4 weeks apart. Observers repeated their readings without knowledge of the previous results. The order of the X-rays was randomized using a sequential random number generator to prevent possible recall. Observers were blinded to the clinical outcome of the patients.

Statistical analysis

A descriptive analysis of the continuous and categorical data was performed using proportions, frequency distributions, means, and standard deviations. K-statistics were used to establish a relative level of agreement between observers for the two readings and between separate readings by the same observer. Interpretation of the data was performed according to Landis and Koch (9). An agreement is graded as slight (kappa = 0–0.2), fair (kappa = 0.21–0.40), moderate (kappa = 0.41–0.60), substantial (kappa = 0.61–0.80) and almost perfect (kappa = 0.81–1). At the end of the study, two observers rated all patients together and declared a consensus. This rating was accepted as the reference for correlation analysis. Correlation between the radiographic grade and clinical status was performed using Pearson correlation analysis. A p value less than 0.05 was set as significant.

Table 1. Kellgren-Lawrence classification (4)

Grade	Kellgren -Lawrence Classifications
0	No radiographic findings of osteoarthritis
1	Minute osteophytes of doubtful clinical significance
2	Definite osteophytes with unimpaired joint space
3	Definite osteophytes with moderate joint space narrowing
4	Definite osteophytes with severe joint space narrowing and subchondral sclerosis

Table 2. Takakura classfications (6)

Grade	Takakura Classification
1	No joint space narrowing but early sclerosis and osteophyte formation
2	Narrowing of the joint space medially
3a	Obliteration of the joint space limited to the facet of medial malleolus with subchondral bone contact
3b	Obliteration of the joint space advanced to the roof of the talar dome with subchondral bone contact
4	Obliteration of the whole joint space with complete bone contact

RESULTS

Intra-observer reliability for KL was fair for both observers (κ: 0.277, and κ: 0340 respectively). Interobserver reliability for KL was fair (κ:0.362) at first occasion, slight (κ:0.106) at the second occasion. Intra-observer reliability of Takakura classification was fair (κ: 0.255) for observer A, and moderate (κ: 0.576) for observer B. Interobserver reliability for Takakura classification was moderate (κ: 0.472) at first occasion, slight (κ: 0.114) at the second occasion. Intra-observer reliability of van Dijk classification was fair (κ:0.321) for observer A, and moderate (κ:0.443) for observer B. Interobserver reliability for van Dijk classification was fair (κ:0.328) at first occasion, slight (κ:0.163) at the second occasion. Summary of results are presented in Table 4. There was a significant correlation between all classification schemes and AOFAS (Table 5). Correlation coefficient (Pearson rho) was higher for Takakura classification compared to KL and van Dijk classifications.

Table 3. Van Dijk classification (5)

Grade	van Dijk classification
0	Normal joint or subchondral sclerosis
1	Osteophytes without joint space narrowing
2	Joint space narrowing with or without osteophytes
3	(Sub)total disappearance or deformation of the joint space

Table 4. Summary of reliability tests.

	Kellgren-Lawrence			Takakura			van Dijk		
	Kappa	95% CI	Agreement (%)	Kappa	95% CI	Agreement (%)	Kappa	95% CI	Agreement (%)
A t_1 vs. A t_2	0.277	0.131-0.422	41.66	0.255	0.072-0.182	51.66	0.321	0.144-0.497	53.33
B t_1 vs. B t_2	0.340	0.193-0.487	46.66	0.576	0.417-0.734	60.00	0.443	0.282-0,603	58.33
A t_1 vs. B t_1	0.362	0.203-0.520	50.00	0.472	0.303-0.640	63.33	0.328	0.171-0.484	50.00
A t_2 vs. B t_2	0.106	-0.039-0.251	30.00	0.114	0.052-0.280	41.66	0.163	0.004-0.321	38.33

Abbreviations A: Observer A, B: Observer B, t1: First rating, t2: second rating, CI: confidence interval.

Table 5. Correlation between ankle OA classifications and AOFAS.

		AOFAS
Kellgren Lawrence	Pearson Correlation (rho)	-0.591
	Sig. (2-tailed)	0.000
Takakura	Pearson Correlation (rho)	-0.642
	Sig. (2-tailed)	0.000
Van Dijk	Pearson Correlation (rho)	-0.547
	Sig. (2-tailed)	0.000

DISCUSSION

In this study, the interobserver and intra-observer reliability of three different classification systems used for radiological evaluation of post traumatic ankle OA and the compatibility of radiological stage with clinical status were investigated. In general, the *kappa* value is expected to be above 0.80 for a classification system to be considered reliable and reproducible (9). Based on our study, all three of these classification systems were observed to have unacceptable inter-observer and intra-observer reliability. However, a negative correlation was observed between the classification stage and clinical status of the patients. In other words, there was a decrease in AOFAS scores in patients with advanced stage ankle OA. The highest kappa values were found in the Takakura classification when all three classification methods were compared. Furthermore, Takakura classification was found to be superior to other classifications in terms of clinical compliance. Therefore, we recommend the use of the Takakura classification system in posttraumatic ankle OA radiological evaluation.

The first radiological classification systems for OA were created by Kellgren and Lawrence in 1957 (4). It has been suggested that this classification system was developed for all joints. In their original study, various joints were classified using the same scheme including distal interphalangeal joint (DIP), metacarpophalangeal joint (MCP), first carpometacarpal (CMC), wrist, cervical and lumbar vertebrae, hip and knee joints. However, ankle joint was not included. Highest interobserver reliability (kappa:0.83) was found in the tibiofemoral part of the knee joint. They recommended evaluation of OA classification should be done with at least two observers with consensus to increase the validity. However, KL classification is still widely used classification system for OA(10). In current literature, there are limited number of studies that investigated the reliability of KL classification for ankle OA. Holzer et al. investigated the reliability of KL classification and its correlation with clinical symptoms in ankle OA. However, they modified the radiological criteria and further subdivided the Grade III ankle OA into Grade IIIA and IIIB according to presence of talar tilt. Moreover, they indicated the location of osteophytes according to involvement of the joint such as medial tibiotalar, superior tibiotalar and talofibular articulations. Thus, they converted subjective original definitions to more objective descriptions. They reported moderate reliability of original KL classification, but the reliability increased to good agreement with their modifications. (8,11). In addition to the ankle tibiotalar joint, a study was performed to analyze the reliability of KL classification for subtalar (ST) and talonavicular (TN) joints. In this study, ST and TN joints were classified according to KL classification in patients who underwent total ankle arthroplasty. The authors found the kappa value 0.43 for the ST joint, and the kappa value 0.37 for the TN joint. Similarly, the authors stated that the KL system was not suitable tool for the classification of subtalar joint. (12).

The other ankle OA classification system is created by Van Dijk et al. performed arthroscopic debridement in 34 patients who developed osteoarthritis and anterior impingement after various ankle trauma (5). They classified all patients in accordance with their own staging system. The functional outcomes were better in patients within Grade 0 and I compared to patients in Grade II and III. They suggested that the use of this classification system might predict the results before ankle arthroscopic interventions. However, they have not performed any reliability studies related to this classification system. The third calcification system is created by Takakura et al (6). Retrospectively evaluated the results of distal tibial valgus osteotomy in varus type ankle OA patients. In contrast to KL and Van Dijk classification systems, they subdivided grade III into grade III A and grade III B according to the extent of obliteration of the joint space. In grade III A, the obliteration was limited to the medial tibiotalar joint, whereas in Grade III B the obliteration involved the whole tibiotalar joint. Satisfactory outcomes were obtained in patients with stage 2 - 3a after valgus osteotomy. However, there was no reliability analysis of their classification in the original article.

In current literature, there are only one study evaluating the reliability of all these classifications (KL, Van Dijk, Takakura) for ankle OA. Claussen et al. sent 128 ankle radiographs (Ap and lateral)

to 118 orthopedic surgeons on a web-based system (7). Observers classified these ankle radiographs according to KL, Takakura and Vandijk classification only once. They found unacceptable interobserver reliability in all classification systems and they recommended no to use these systems in clinical decision making. The strongest aspect of this study was inclusion of several observers with different level of experience, and the study was performed on high number of radiographs. However, the radiographs were sent over the internet and the evaluations were not performed in the same environment which may have caused a lack of standardization. Moreover, only interobserver reliability analysis was performed. Our study is the first study which presents both inter and intra-observer reliability analysis in current literature. The results of our study are compatible with these previous few studies.

Our study has some weak and strong aspects. The number of cases were relatively small compared to similar studies in current literature. All cases were operatively treated ankle fractures, thus primary ankle OA cases were not evaluated. Both observers were experienced surgeons who were using these classification systems in their routine practice. Therefore, the effect of level of experience could not be realized. On the other hand, final clinical status of the patients was compared with the radiological stage of the disease. Both interobserver and intra-observer reliability analysis was performed.

In conclusion, the widely used KL, Van Dijk and Takakura classifications for ankle OA have been found to be below acceptable thresholds regarding reliability. These classifications are based on subjective definitions. It is obvious that, a new and more objective ankle OA classification system is necessary. In order to increase the reliability and reproducibility of an ankle OA classification, a specific system to ankle joint should be constituted. In addition, more objective and numerical values should be included as classification criteria. With the introduction of digital radiology in daily practice, it is possible to specify joint space measurements or the degree of subchondral sclerosis by numerical data. In future, a quantitative classification instead of a qualitative classification should be developed.

Acknowledgements

We would like to thank to Dr. Özkan Köse for his help on the statistical analysis and preparation of tables.

REFERENCES

1. Barg A, Pagenstert GI, Hügle T, Gloyer M, Wiewiorski M, Henninger HB et al. Ankle Osteoarthritis. Foot Ankle Clin 2013; 18(3): 411-26.
2. Cheng YM, Huang PJ, Hung SH, Chen TB, Lin SY. The surgical treatment for degenerative disease of the ankle. Int Orthop 2000; 24(1): 36-9.
3. Ewalefo SO, Dombrowski M, Hirase T, Rocha JL, Weaver M, Kline A et al. Management of posttraumatic ankle arthritis: literature review. Curr Rev Musculoskelet Med 2018;11(4):546-57.
4. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957; 16(4): 494-502.
5. van Dijk CN, Verhagen RA, Tol JL. Arthroscopy for problems after ankle fracture. J Bone Joint Surg Br 1997; 79(2):280-4.
6. Tanaka Y, Takakura Y, Hayashi K, Taniguchi A, Kumai T, Sugimoto K. Low tibial osteotomy for varus-type osteoarthritis of the ankle. J Bone Joint Surg Br 2006; 88 (7): 909-13.
7. Claessen FMAP, Meijer DT, van den Bekerom MPJ, Deynoot BDJG, Mallee WH, Doornberg JN et al. Reliability of classification for post-traumatic ankle osteoarthritis. Knee Surgery Sport Traumatol Arthrosc 2016; 24(4): 1332-7.
8. Holzer N, Salvo D, Marijnissen ACA, Vincent KL, Ahmad AC, Serra E et al. Radiographic evaluation of posttraumatic osteoarthritis of the ankle: the Kellgren–Lawrence scale is reliable and correlates with clinical symptoms. Osteoarthr Cartil 2015; 23(3): 363-9.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33(1): 159-74.
10. Kohn MD, Sassoon AA, Fernando ND. Classifications in Brief: Kellgren-Lawrence Classification of Osteoarthritis. Clin Orthop Relat Res 2016; 474(8): 1886-93.

11. Moon J-S, Shim J-C, Suh J-S, Lee W-C. Radiographic predictability of cartilage damage in medial ankle osteoarthritis. Clin Orthop Relat Res 2010; 468(8): 2188-97.

12. Mayich DJ, Pinsker E, Mayich MS, Mak W, Daniels TR. An Analysis of the Use of the Kellgren and Lawrence Grading System to Evaluate Peritalar Arthritis Following Total Ankle Arthroplasty. Foot Ankle Int 2013; 34(11): 1508-15.

Corresponding Author

Adil TURAN (Başasistan)
Sağlık Bilimleri Üniversitesi, Antalya Eğitim ve Araştırma Hastanesi, Ortopedi ve Travmatoloji Kliniği, Antalya
Phone: +90 5052381386
E-mail: adilturan@yahoo.com
ORCID: 0000-0002-8062-2330

Cemil AKTAN (Op. Dr.)ORCID:0000-0002-8245-2187