

Intraobserver and Interobserver Reliability of Three Classification Systems for Hallux Rigidus

Peter Pham, DPM*
 Sarita Dillard, DPM*
 Christina Schilero, DPM*
 Sharon Chiang, PhD†

Background: There are more than ten classification systems currently used in the staging of hallux rigidus. This results in confusion and inconsistency with radiographic interpretation and treatment. The reliability of hallux rigidus classification systems has not yet been tested. We sought to evaluate the intraobserver and interobserver reliabilities of three commonly used classifications for hallux rigidus.

Methods: Twenty-one plain radiograph sets were presented to ten American College of Foot and Ankle Surgeons board-certified foot and ankle surgeons. Each physician classified each radiograph based on clinical experience and knowledge according to the Regnauld, Roukis, and Hatstrup and Johnson classification systems. The two-way mixed single-measure consistency intraclass correlation coefficient was used to calculate intrarater and interrater reliabilities.

Results: The mean \pm SD intrarater reliability of individual sets for the Roukis (0.62 ± 0.19) and Hatstrup and Johnson (0.62 ± 0.28) classification systems was fair to good and for the Regnauld system bordered between fair to good and poor (0.43 ± 0.24). The interrater reliability of the mean classification was excellent for all three classification systems.

Conclusions: Reliable and reproducible classification systems are essential for treatment and prognostic implications in hallux rigidus. Herein, the Roukis classification system had the best intrarater reliability. Although there are various classification systems for hallux rigidus, the present results indicate that the three classification systems evaluated show reliability and reproducibility. (J Am Podiatr Med Assoc 110(3): 1-5, 2020)

Hallux limitus and hallux rigidus are two of the most common pathologic disorders facing podiatric physicians, affecting 2.5% of the adult population.¹ Hallux limitus is an arthritic condition causing decreased motion at the first metatarsophalangeal joint, which tends to be more frequent in men and with 44% of affected patients older than 80 years. The decrease in sagittal plane motion leads to pain, stiffness, and gait disturbances.^{2,3} The condition is progressive and oftentimes results in hallux rigidus, a complete lack of motion at the first metatarsophalangeal joint. Although there does not seem to be one underlying cause of hallux limitus, it can be

thought to result from hypermobility of the first ray, pes planus, a long first metatarsal, primus elevatus, iatrogenous complications, inflammatory diseases such as gout and rheumatoid arthritis, and genetics.^{2,4,5}

Hallux rigidus and hallux limitus can be diagnosed with radiographs and clinical evaluation.^{2,6,7} The degree of joint space narrowing and the presence of osteophytes and first metatarsal elevation help determine the severity of the disorder. There have been numerous radiographic classification systems to help measure and grade hallux rigidus.⁸⁻¹⁴ The three most commonly used classifications are the Regnauld, Hatstrup and Johnson, and Roukis classification systems. The Regnauld classification has three grades, with grade I defined as functional hallux limitus, grade II as joint adaptation with flattening of the first metatarsal head and pain

*Department of Podiatry, West Houston Medical Center, Houston, TX.

†Baylor College of Medicine, Houston, TX.

Corresponding author: Peter Pham, DPM, West Houston Medical Center, 12121 Richmond Ave, Suite 415, Houston, TX 77082. (E-mail: peter.duypham@gmail.com)

at end range of motion, and grade III as arthrosis with severe flattening of the first metatarsal head, osteophytes, asymmetrical joint space narrowing, and erosions.^{8,11} The Hatstrup and Johnson classification also has three grades: grade I is characterized as mild-to-moderate formation of osteophytes with no joint space involvement; grade II as moderate osteophyte formation, joint space narrowing, and subchondral sclerosis; and grade III as increased osteophyte formation and loss of joint space.⁹ Roukis was the first grading system applied prospectively and is similar to the Regnault classification but includes a stage IV, which is defined as having less than 10° of range of motion and loose bodies with obliteration of joint space in the first metatarsophalangeal joint.¹⁰

Surgeons rely heavily on radiographs and classification systems to confirm clinical diagnosis and aid in surgical planning.⁶ The objective of this study was to assess the interobserver and intraobserver reliability of the Regnault, Hatstrup and Johnson, and Roukis classification systems of hallux rigidus. In particular, we examine the consistency of three commonly used classification systems and discuss implications for the treatment of hallux rigidus.

Materials and Methods

We selected the Regnault, Roukis, and Hatstrup and Johnson hallux rigidus classification systems for study (Table 1). Twenty-one plain radiograph sets (three radiograph packets with each packet containing seven sets of radiographs) of the foot were randomly selected from physicians' electronic systems based on the presence of hallux limitus deformity. Three standard views were selected and provided per set (weightbearing anteroposterior,

medial oblique, and lateral) (Fig. 1). Each radiograph packet contained the same seven sets of radiographs with one of the three classification systems attached.

Ten American College of Foot and Ankle Surgeons (ACFAS) board-certified foot and ankle surgeons were instructed to objectively classify each packet of radiographs according to its assigned classification system based on their clinical experience and knowledge. The surgeons were randomly selected. A handout of the instructions was given to each physician. This process was repeated with each individual physician three times to determine whether each physician classified each set of radiographs consistently and to eliminate potential bias. All of the surgeons were familiar with the Regnault, Roukis, and Hatstrup and Johnson classification systems (Table 1).

Statistics

Physician intrarater and interrater reliabilities were assessed based on the two-way mixed single-measure consistency intraclass correlation coefficient (ICC).^{14,15} The ICC provides a measure of the proportion of reliable variance and typically ranges between 0 and 1.¹⁵ Specifically, ICC[3,1] and ICC[3,k] are reported. The ICC was interpreted based on guidelines for reliability provided by Fleiss¹⁶: less than 0.40, poor reliability; 0.40 to 0.75, fair to good reliability; and greater than 0.75, excellent reliability. Analyses were conducted using R version 3.1.3 (Vienna, Austria).¹²⁻¹⁴

Results

The 21 radiograph sets were assessed by the ten ACFAS board-certified foot and ankle surgeons:

Table 1. Comparison of the Three Hallux Rigidus Classification Systems Used in the Study¹⁰⁻¹²

Classification	Grade I	Grade II	Grade III	Grade IV
Regnault	Mild limitation of dorsiflexion, mild dorsal spurring, pain, subchondral sclerosis	Flattening, broadening of metatarsal head and base of proximal phalanx, sesamoid hypertrophy	Worsening loss of joint space, ankylosis, extensive osteophytes	
Roukis	Metatarsus primus elevatus, periarticular subchondral sclerosis, minimal dorsal exostosis, and minimal flattening of metatarsal head	Moderate dorsal exostosis, flattening of metatarsal head, minimal joint space narrowing, sesamoid hypertrophy	Severe dorsal exostosis, focal joint space narrowing, cyst formation, loose bodies	Excessive exostosis of metatarsal head and proximal phalanx base, absent joint space, ankylosis
Hatstrup and Johnson	Mild-to-moderate osteophyte formation, joint space normal	Moderate osteophyte formation, joint space narrowing, subchondral sclerosis	Marked osteophyte formation, loss of visible joint space with or without subchondral cyst formation	



Figure 1. Examples of radiographs used in this study: weightbearing anteroposterior (A) and lateral/medial oblique (B) views of the left foot.

nine men and one woman with 6 to 7 years of postgraduate education and 5 to more than 20 years of experience. Overall, the mean \pm SD intrarater reliability of individual sets for the Roukis and Hatstrup and Johnson classification systems was fair to good (0.62 ± 0.19 and 0.62 ± 0.28 , respectively), whereas that for the Regnaud system bordered between fair to good and poor (0.43 ± 0.24). The mean \pm SD intrarater reliability of the mean classification across the three sets was excellent for the Roukis and Hatstrup and Johnson classification systems (0.81 ± 0.13 and 0.78 ± 0.23 , respectively) and fair to good for the Regnaud system (0.65 ± 0.24). Intrarater and interrater reliabilities are shown in Table 2.

Interrater reliability was lower for set 3 than for sets 1 and 2 for all classification systems (Table 2).

The Roukis classification system attained consistently higher interrater reliability than the Regnaud or Hatstrup and Johnson classification systems across sets 1, 2, and 3. Specifically, interrater reliability of individual physicians for the Roukis classification system was fair to good, whereas that for the Regnaud and Hatstrup and Johnson classification systems was fair to good for sets 1 and 2 and poor for set 3. Reasons for the possible outliers in set 3 could be attributed to the subjective reading as well as variability in radiographic views and the different professional backgrounds and experience of the physicians. Interrater reliability of the mean classification across the ten physicians was excellent for all three classification systems, with the highest reliability attained by the Roukis classification system.

Table 2. Physician Intrarater and Interrater Reliabilities for the Three Studied Classification Systems for Hallux Rigidus

Reliability	Regnaud	Roukis	Hatstrup and Johnson
ICC for intrarater reliability, mean \pm SD			
ICC[3,1]	0.43 ± 0.24	0.62 ± 0.19	0.62 ± 0.28
ICC[3,k]	0.65 ± 0.24	0.81 ± 0.13	0.78 ± 0.23
ICC for interrater reliability			
ICC[3,1], set 1	0.53	0.66	0.53
ICC[3,1], set 2	0.54	0.60	0.54
ICC[3,1], set 3	0.29	0.40	0.29
ICC[3,k], set 1	0.90	0.94	0.90
ICC[3,k], set 2	0.90	0.92	0.90
ICC[3,k], set 3	0.76	0.84	0.76

Abbreviation: ICC, intraclass correlation coefficient.

Discussion

We evaluated the intrarater and interrater reliability of three commonly used hallux rigidus classification systems in evaluating foot radiographs in a sample of ten ACFAS board-certified surgeons. In doing so, we evaluated whether particular classification systems may exhibit greater consistency for the grading of hallux rigidus. The present results show consistency and little variability in the Regnauld, Roukis, and Hatstrup and Johnson classification systems.

Hallux rigidus of the first metatarsophalangeal joint is one of the most common conditions associated with the foot. A multitude of classification systems for hallux rigidus have been used since 1930.¹⁷ Previous research has discussed the relative faults and strengths of the various classification systems, including reliance solely on radiographic findings without accounting for clinical findings.¹⁷ To our knowledge, we found only one other study, by Pate et al,¹⁸ that compares the reliability and intraobserver agreement of these classification systems. They suggested using radiographic grading systems for hallux rigidus with caution because they had only 75% intraobserver reliability. In the present study, we found that the Roukis and Hatstrup and Johnson classification systems obtain higher intrarater reliability than the Regnauld system. With respect to interrater reliability, the Roukis classification system attains higher interrater reliability than the Regnauld and Hatstrup and Johnson classification systems.

The classification systems evaluated herein are useful for grading the severity of hallux rigidus and for directing treatment or evaluating prognosis. For example, a Roukis grade I would lead to either conservative treatment or a cheilectomy, Watermann, or Youngswick-Austin procedure, and a Roukis grade IV would lead more to an implant versus arthrodesis.^{12,19} Conservative treatments for hallux rigidus include icing, nonsteroidal therapy, shoe modification, corticosteroid injections, physical therapy, and orthotic devices.¹⁸ The goal for surgical treatment is to relieve pain and ensure mobility of the first metatarsophalangeal joint. The present study offers the first quantitative comparison of the relative strengths of the classification systems based on interrater and intrarater reliability. This information can be useful in determining the plan and treatment that an individual needs based on the level of severity.

Limitations of this study include the relatively small number of observers and radiographs used.

Future research using larger sample sizes may be useful to validate the conclusions identified herein. In addition, in the present study we focus on the Roukis, Regnauld, and Hatstrup and Johnson classification systems. These particular systems were chosen due to their widespread use. However, other classification systems, such as Hanft, Coughlin, Drag, Oloff, and Jacobs, may be useful to be evaluated in context.^{20,21}

The present data show consistent findings of reliability and reproducibility using the three classification systems, with the Roukis classification having the best intrarater reliability. Currently, there is no gold standard among hallux rigidus classification systems, resulting in difficulty comparing results from different classification systems from a physician perspective.^{6,21} We conclude that the classification system for hallux rigidus is a reliable and reproducible tool in evaluating radiographs and grading hallux rigidus. Larger randomized prospective studies may be useful to evaluate the conclusions observed herein.

Financial Disclosure: None reported.

Conflict of Interest: None reported.

References

1. KEISERMAN LS, SAMMARCO VJ, SAMMARCO GJ: Surgical treatment of hallux rigidus. *Foot Ankle Clin* **10**: 75, 2005.
2. COUGHLIN MJ, SHURNAS PS: Hallux rigidus: demographics, etiology, and radiographic assessment. *Foot Ankle Int* **24**: 731, 2003.
3. BOTEK G, ANDERSON MA: Etiology, pathophysiology and staging of hallux rigidus. *Clin Podiatr Med Surg* **28**: 229, 2011.
4. YEE G, LAU J: Current concepts review: hallux rigidus. *Foot Ankle Int* **29**: 392, 2008.
5. CHANG TJ: Stepwise approach to hallux limitus: a surgical perspective. *Clin Podiatr Med Surg* **13**: 449, 1996.
6. ZGONIS T, JOLLY GP, GARBALOSA JC: The value of radiographic parameters in the surgical treatment of hallux rigidus. *J Foot Ankle Surg* **44**: 184, 2005.
7. SHEREFF MJ, BAUMHAUER JF: Hallux rigidus and osteoarthritis of first metatarsophalangeal joint. *J Bone Joint Surg Am* **80**: 898, 1998.
8. VANORE JV, CHRISTENSEN JC, KRAVITZ SR, ET AL: Diagnosis and treatment of first metatarsophalangeal joint disorders, section 1: hallux rigidus. *J Foot Ankle Surg* **42**: 112, 2003.
9. DRAGO JJ, OLOFF L, JACOBS AM: A comprehensive review of hallux limitus. *J Foot Surg* **23**: 213, 1984.
10. HATSTRUP SJ, JOHNSON KA: Subjective results of hallux rigidus following treatment with cheilectomy. *Clin Orthop Relat Res* **226**: 182, 1988.

11. REGNAULD B: *The Foot: Pathology, Aetiology, Semiology, Clinical Investigation and Therapy*, Springer-Verlag, New York, 1986.
12. ROUKIS TS, JACOBS M, DAWSON DM, ET AL: A prospective comparison of clinical radiographic, and intraoperative features of hallux rigidus. *J Foot Ankle Surg* **41**: 76, 2002.
13. COUGHLIN MJ, SHURNAS PS: Hallux rigidus: grading and long-term results of operative treatment. *J Bone Joint Surg Am* **85**: 2072, 2003.
14. NORMAN GR, STREINER DL: *Biostatistics: The Bare Essentials*, BC Decker, Hamilton, ON, Canada, 2008.
15. SHROUT PE, FLEISS JL: Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* **86**: 420, 1979.
16. FLEISS J: *The Design and Analysis of Clinical Experiments*, John Wiley & Sons Inc, New York, 1986.
17. BEESON P, PHILLIPS C, COOR S: Classification system for hallux rigidus: a review of the literature. *Foot Ankle Int* **29**: 407, 2008.
18. PATE RC, FANNING JW, SHIELDS NN, ET AL: Reliability of hallux rigidus radiographic grading system. *Kansas J Med* **8**: 125, 2015.
19. TARANOW WS, MOORE J: Hallux rigidus: a treatment algorithm. *Tech Foot Ankle* **11**: 65, 2012.
20. SHURNAS PS: Hallux rigidus: etiology, biomechanics and nonoperative treatment. *Foot Ankle Clin* **14**: 1, 2009.
21. HANFT JR, MASON ET, LANDSMAN AS, ET AL: A new radiographic classification for hallux limitus. *J Foot Ankle Surg* **32**: 397, 1993.