Intra- and Interobserver Reliability of Three Classification Systems for Hallux Rigidus

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

*West Houston Medical Center, Houston, TX.

†Baylor College of Medicine, Houston, TX.

Abstract

Background: There are over 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. The purpose of this study was to evaluate intra- and interobserver reliability using three commonly used classifications for hallux rigidus.

Methods: Twenty-one plain radiograph sets were presented to ten ACFAS board-certified foot and ankle surgeons. Each physician classified each radiograph based on clinical experience and knowledge according to the Regnault, Roukis, and Hattrup and Johnson classification systems. The two-way mixed single-measure consistency intraclass correlation was used to calculate intra- and interrater reliability.

Results: The intrarater reliability of individual sets for the Roukis and Hattrup and Johnson classification systems was “fair to good” (Roukis, 0.62 ± 0.19 ; Hattrup and Johnson, 0.62 ± 0.28), whereas the intrarater reliability of individual sets for the Regnault 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. In our study, Roukis classification system had the best intrarater reliability. Although there are various classification systems for hallux rigidus, our results indicate that all three of these classification systems show reliability and reproducibility.

Hallux limitus and hallux rigidus are two of the most common pathologies facing podiatrists, 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 over the age of 80. 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 to

Corresponding author: Peter Pham, DPM, 12121 Richmond Avenue, Ste 415, Houston, TX 77082. (peter.duypham@gmail.com).

Financial Disclosure: None reported.

Conflict of Interest: None reported.

hallux limitus, it can be thought to result from hypermobility of the first ray, pes planus, long first metatarsal, primus elevates, forefoot, iatrogenous complications, inflammatory diseases such as gout and rheumatoid as well as genetics.^{2,4,5}

Both hallux rigidus and hallux limitus can be diagnosed both through radiographs and clinical evaluation.^{2,6,7} The degree of joint space narrowing, and presence of osteophytes and first metatarsal elevation help determine the severity of the disorder. There have been numerous radiographic classifications systems to help measure and grade hallux rigidus.⁸⁻¹⁴ The three most commonly used classifications are Regnault, Hattrup and Johnson, and Roukis classification systems. The Regnault classification has three grades, with grade I defined as functional hallux limitus; grade II defined as joint adaptation with flattening of the first metatarsal head and pain at end range of motion; and grade III defined as arthrosis with severe flattening of the first metatarsal head, osteophytes, asymmetric joint space narrowing, and erosions.^{8,11} Hattrup 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; 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 degrees of range of motion, loose bodies with obliteration of joint space within the first metatarsophalangeal joint.¹⁰

Surgeons rely heavily on radiographs and classification systems to both confirm clinical diagnosis as well as aid in surgical planning.⁶ The objective of this study is to look at the inter-and intra-observer reliability for the Regnault, Hattrup 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 following classification systems for hallux rigidus: Regnault, Roukis, and Hattrup and Johnson classifications (Table 1). Twenty-one plain film radiograph sets (three radiograph packets with each packet containing seven sets of radiographs) of the foot were randomly selected from physicians electronic base system based on presence of hallux limitus deformity. Three standard views were selected and provided per set (weight-bearing anterior-posterior, medial oblique, lateral (Fig 1A, 1B)). Each radiograph packet contained the same seven sets of radiographs with one of the above classification systems attached.

Ten 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 all randomly selected (Table 2). A handout of the instructions was given to each physician. This process was repeated with each individual physician a total of three times, in order to determine whether each physician classified each set of radiographs consistently and to eliminate potential bias. All surgeons

were familiar with the Regnauld, Roukis, and Hattrup and Johnson classification systems (Table 1).

Statistics

Physician intra- and interrater reliabilities were assessed based on the two-way mixed single-measure consistency intraclass correlation (ICC) (Norman and Streiner, 2008; Shrout and Fleiss, 1979). [13] The ICC provides a measure of the proportion of reliable variance and typically ranges between 0 and 1 (Shrout and Fleiss, 1979).⁴ Specifically, ICC(3,1) and ICC(3,k) were reported. The ICC was interpreted based on guidelines for reliability provided by (Fleiss, 1986): <0.40, “poor” reliability; 0.40–0.75, “fair” to “good” reliability; >0.75, “excellent” reliability. Analyses were conducted using R version 3.1.3 (Vienna, Austria).
12–14

RESULTS

The twenty-one radiograph sets were assessed by the ten ACFAS board-certified foot and ankle surgeons (surgeon experiences provided in Table 2). Overall, the intrarater reliability of individual sets for the Roukis and Hattrup and Johnson classification systems was “fair to good” (Roukis, 0.62 ± 0.19 ; Hattrup and Johnson, 0.62 ± 0.28), whereas the intrarater reliability of individual sets for the Regnauld system bordered between “fair to good” and “poor” (0.43 ± 0.24). Intrarater reliability of the mean classification across the three sets was “excellent” for the Roukis and Hattrup and Johnson classification systems (Roukis, 0.62 ± 0.19 ; Hattrup and Johnson, 0.62 ± 0.28) and “fair to good” for the Regnauld system (0.65 ± 0.24). Intra- and interrater reliabilities are shown in Table 3.

Interrater reliability was lower for Set 3 than for Sets 1 and 2 for all classification systems. The Roukis classification system attained consistently higher interrater reliability than the Regnauld or Hattrup and Johnson classification systems across Sets 1, 2, and 3. Specifically, interrater reliability of individual physicians for the Roukis classification systems was “fair to good,” whereas the interrater reliability of individual physicians for the Regnauld and Hattrup 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 x-ray views and different professional backgrounds and experience from 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.

DISCUSSION

In this study, we evaluate the intra- 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 evaluate whether particular classification systems may exhibit greater consistency for the grading of hallux rigidus. Our results show consistency and little variability in the Regnauld, Roukis, and Hattrup 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 among 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 Patel et al. that compares the reliability and intraobserver agreement of these classification systems.¹⁹ They suggested using radiographic grading system for hallux rigidus with caution as they only had 75% intraobserver reliability. In our study, we found that the Roukis and Hattrup and Johnson classification systems obtain higher intrarater reliability than the Regnault system. With respect to interrater reliability, the Roukis classification system attains higher interrater reliability than both the Regnault and Hattrup & Johnson classification systems.

The classification systems evaluated in our study 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 cheilectomy, waterman or youngswick Austin, while a Roukis grade IV would lead more to an implant versus arthrodesis.^{12,18} Conservative treatments for hallux rigidus include icing, non-steroidal therapy, shoe modification, corticosteroid injections, physical therapy and orthotic therapy.¹⁸ The goal for surgical treatment is to relieve pain and ensure mobility of the first metatarsophalangeal joint. Our study offers the first quantitative comparison of the relative strengths of the classification systems based on inter- and intrarater reliability. This information can be useful in determining the plan and treatment an individual needs based on their level of severity.

A few limitations to our study include the relatively small number of observers and radiographs used. Future research employing larger sample sizes may be useful to validate the conclusions identified here. Additionally, in our study, we focus on the Roukis, Regnault, and Hattrup and Johnson classification systems. These particular systems were chosen due to their widespread usage. However, other classification systems, such as Hanft, Coughlin, Drag, Oloff, and Jacobs may be useful to be evaluated in context.^{20,21}

According to our data, we show consistent findings showing 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 system, 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. A larger randomized prospective studies may be useful to evaluate the conclusions observed here.

REFERENCES

1. Keiserman L, Sammarco J, Sammarco GJ: Surgical treatment of hallux rigidus. *Foot Ankle Clin N Am.* 10: 75–96, 2005.
2. Coughlin MJ, Shurnas PS: Hallux rigidus: demographics, etiology, and radiographic assessment. *Foot Ankle Int.* 24:731–743, 2003. [PubMed: 14587987]

3. Botek G, Anderson MA Etiology, pathophysiology and staging of hallux rigidus. *Clin Podiatry Med Surg.* 28:229–243, 2011.
4. Yee G, Lau J: Current concepts review: hallux rigidus. *Foot Ankle Int.* 29:392–399, 2008.
5. Chang TJ: Stepwise approach to hallux limitus: a surgical perspective. *Clin. Podiatr. Med. Surg* 13:449–459, 1996. [PubMed: 8829035]
6. Zgonis T, Jolly GP, Garbalosa JC: The value of radiographic paramters in the surgical treatment of hallux rigidus. *J Foot Ankle Surg.* 44:184–189, 2005. [PubMed: 15940596]
7. Mj Shereff, Baumhauer JF: Hallux rigidus and osteoarthritis of first metatarsophalangeal joint. *J Bone Joint Surg Am.* 80:898–908. 1998. [PubMed: 9655109]
8. Vanore JV, Christensen JC, Kravitz SR, Schubert JM, Thomas JL, Weil LS, Zlotoff HJ, Mendicino RW, Couture SD: Clinical Practice Guideline First Metatarsophalangeal Joint Disorders Panel of the American College of Foot and Ankle Surgeons. *J Foot Ankle Surg.* 2003.
9. Drago JJ, Oloff L, Jacobs AM: A comprehensive review of hallux limitus. *J. Foot Surg.* 23:213–220, 1984. [PubMed: 6376607]
10. Hattrup SJ, Johnson KA: Subjective results of hallux rigidus following treatment with cheilectomy. *Clin Orthop Relat Res.* 226:182–191, 1988.
11. Ragnauld B: The foot: pathology, aetiology, seminology, clinical investigation and therapy. New York: Springer-Verlag, 1986.
12. Roukis TS, Jacobs M, Dawson DM, Erdmann BB, Ringstrom JB: A prospective comparison of clinical radiographic, and intraoperative features of hallux rigidus. *J Foot Ankle Surg.* 41:76–95, 2002. [PubMed: 11995839]
13. Coughlin MJ, Shurnas PS: Hallux rigidus: grading and long-term results of operative treatment. *J Bone Joint Surg Am.* 85-A:2072–2088, 2003.
14. Norman GR, Streiner DL: Biostatistics: The bare essentials. BC Decker, Hamilton, Ontario, 2008.
15. Shrout PE, Fleiss JL Intraclass correlations: Uses in assessing rater reliability. *Psychological bulletin* 86 420–428, 1979. [PubMed: 18839484]
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.* 29:407–414, 2008. [PubMed: 18442456]
18. Taranow WS, Moore J: Hallux rigidus: A treatment algorithm. *Tech Foot and ankle*;11:65–73, 2012.
19. Pate RC, Fanning J, Shields N, Chong A: Reliability of hallux rigidus radiographic grading system. *Kansas Journal of Medicine.* 125–134, 2015.
20. Shurnas PS. Hallux rigidus: Etiology, biomechanics and nonoperative treatment. *Foot Ankle Clin.* 14: 1–8, 2009. [PubMed: 19232987]
21. Hanft JR, Mason ET, Landsman AS, Kashuk KB: A new radiographic classification for hallux limitus. *J Foot Ankle Surg.* 32: 397–404, 1993. [PubMed: 8251995]
22. Mann RA, Coughlin Mj, Duvries HL. Hallux rigidus: A review of the literature and a method of tratment; *Clin Orthop Relat Res.* 142:57–63, 1979.

Figure 1.
Example of radiographs employed in study (AP/Lateral/MO view of left foot).

Table 1

The three hallux rigidus classification systems used in our study.^{9–11}

Classification	Grade 1	Grade 2	Grade 3	Grade 4
Regnauld	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 elevates, periarticular subchondral sclerosis, minimal dorsal exostosis and minimal flattening of the metatarsal head	Moderate dorsal exostosis, flattening 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
Hattrup 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	

Table 2

Initials of surgeon, gender, years of post-graduate education and experience among surgeons in study sample.

Surgeon	Gender	Years of post grad education	Years of experience
Dr. SM	M	6	20+
Dr. AL	F	6	15+
Dr. JM	M	7	5+
Dr. DB	M	6	20+
Dr. BL	M	6	10+
Dr. SN	M	6	15+
Dr. RB	M	7	10+
Dr. MM	M	7	10+
Dr. MI	M	7	8+
Dr. JH	M	6	15+

Table 3

Physician intrarater and interrater reliability for Regnauld, Roukis, and Hattrup and Johnson classification systems for hallux rigidus.

	Regnauld	Roukis	Hattrup & Johnson
ICC for intrarater reliability, mean (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