



REVISTA DEL PIE Y TOBILLO

www.elsevier.es/rptob



ORIGINAL

Clasificaciones de hallux rigidus: ¿todos vemos lo mismo?

Daniel Chaverri Fierro^{a,*}, Rafael Hernandez Hernandez^a, Elena Gaviria Parada^a,
Alberto Ruiz Nasarre^a, Raúl Plaza García^a, Fernando Alvarez Goenaga^a
y Ramón Viladot Pericé^b

^a Servicio de Cirugía Ortopédica, Hospital Sant Rafael, Barcelona, España

^b Servicio de Cirugía Ortopédica, Clínica Tres Torres, Barcelona, España

Recibido el 13 de junio de 2014; aceptado el 10 de enero de 2015

PALABRAS CLAVE

Hallux rigidus;
Clasificación;
Fiabilidad;
Reproductibilidad

Resumen

Introducción: El hallux rigidus (HR) es una de las causas más frecuentes de dolor en el antepié. Podemos encontrar hasta 18 clasificaciones diferentes para evaluar dicha patología sin existir a día de hoy un consenso claro sobre cual utilizar, si bien la más extendida sería la propuesta por Coughlin y Shurnas. Haciendo una revisión crítica de la literatura observamos que todas ellas han sido elaboradas en base a parámetros basados en la experiencia y opiniones de expertos careciendo todas ellas de estudios de validez y fiabilidad inter/intraobservador.

Objetivos: Valorar la fiabilidad inter/intraobservador de la clasificación de Coughlin & Shurnas en su apartado radiológico.

Material y métodos: 6 traumatólogos (2 expertos, 2 sénior, 2 noveles) analizaron las radiografías (F y P en carga) de 70 pacientes afectados de HR y las catalogaron en base a la gradación de la clasificación de Coughlin y Shurnas. Tras aleatorizar la muestra se repitió la observación a los 3 días de la primera. Se analizó la fiabilidad inter-intraobservador mediante el estadístico Fleiss Kappa y el Coeficiente de Concordancia de Kendall usando el paquete estadístico Minitab 16.1.1

Resultados: Se obtuvo una fiabilidad interobservador de Fleiss Kappa 0,46 y CC Kendall de 0,8183. En cuanto a la consistencia interna o fiabilidad intraobservador esta osciló entre valores Fleiss Kappa 0,52 y 0,84 para los diferentes observadores.

Conclusiones: La clasificación de Coughlin y Shurnas en su apartado radiológico presenta una buena fiabilidad intraobservador, así como una moderada concordancia interobservador en base a nuestros resultados.

© 2015, SEMCPT. Publicado por Elsevier España, S.L.U. Este es un artículo Open Acces distribuido bajo los términos de la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*Autor para correspondencia.

Correo electrónico: danichaverrifierro@gmail.com (D. Chaverri Fierro).

KEYWORDS

Hallux rigidus;
Classification;
Reliability;
Reproducibility
of results

Classification systems for hallux rigidus: See the same thing?**Abstract**

Introduction: Hallux rigidus (HR) is one of the most common causes of pain in the forefoot. We can find 18 different classifications to evaluate this disease without exist nowadays a clear consensus on which to use, although the most widespread it would be Coughlin & Shurnas classification. Making a critical review of the literature we can see that all of them have been developed based on the experience and expert opinion, existing a lack of validity and inter / intra observer reliability studies.

Objectives: To assess the inter / intraobserver reliability of the classification of Coughlin & Shurnas in their radiological paragraph.

Material and Methods: 6 orthopedic surgeons (2 experts, 2 seniors, 2 novice) analyzed the radiographs (F and P load) of 70 patients with HR and categorized them based on the Coughlin & Shurnas classification. After randomizing the sample, the observation was repeated 3 days after the first. Inter-intraobserver reliability was analyzed using Fleiss Kappa statistic and Kendall Coefficient of Concordance using the statistical package Minitab 16.1.1.

Results: A Fleiss Kappa interobserver reliability of 0.46 and CC Kendall of 0.8183 was obtained. Regarding the internal consistency or intraobserver reliability Fleiss Kappa values ranged from 0.52 and 0.84 for different observers.

Conclusions: The classification of Coughlin & Shurnas in their radiological paragraph has good intraobserver reliability and a moderate interobserver agreement based on our results.

© 2015, SEMCPT. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

El hallux rigidus (HR) es una de las causas más frecuentes de dolor en el antepié. En la literatura podemos encontrar descritas hasta 18 clasificaciones diferentes para evaluar esta patología sin existir a día de hoy un consenso claro sobre cual utilizar, si bien la más extendida sería la propuesta por Coughlin y Shurnas^{1,2}.

Cualquier sistema de clasificación debe estar diseñado buscando cumplir los siguientes requisitos:

Validez

- *De contenido:* Se refiere a si realmente el instrumento o escala mide lo que quiere medir.
- *De criterio:* Hace referencia a si los elementos incluidos en el sistema de clasificación se correlacionan con alguna medida considerada como «gold standard».
- *De constructo:* Los elementos incluidos en el sistema de clasificación tienen lugar o pertenecen a una teoría o esquema teórico.

Fiabilidad

- *Reproductibilidad:* Concordancia interobservador
- *Consistencia interna:* Concordancia intraobservador en el test-retest.

Sensibilidad

- Mide la capacidad de cada componente de predecir cambios en el status del paciente. Ejemplo: ¿Un cambio radiológico se correlaciona con el grado de dolor del paciente?

Aplicabilidad clínica

- Las medidas incluidas son aceptables para el paciente en términos de riesgos y para el clínico en lo que respecta a inversión de tiempo de esfuerzo.

Beeson et al.³ realizaron una revisión crítica de la literatura buscando identificar la presencia o no de estos criterios para cada una de las 18 clasificaciones existentes. En su revisión encontraron una falta de consistencia en la construcción de las diferentes clasificaciones, no hallando ningún estudio o artículo que demostrara la validez, fiabilidad, sensibilidad o aplicabilidad clínica de cada una de ellas. No encontraron tampoco ningún estudio ideado en sí mismo para medir o validar una clasificación determinada, sino que la mayoría fueron artículos en los que se utilizaban las escalas para testar o indicar intervenciones.

Todo lo anterior podría, según Beeson et al., invalidar los resultados de investigaciones que se basan en estas clasificaciones, así como las comparaciones entre estudios que usen la misma o diferentes clasificaciones.

A pesar de ello Besson et al. consideran la clasificación de Coughlin y Shurnas como la que más se aproxima al «gold standard». Esta clasificación se basa en parámetros clínico-radiológicos tal y como describe la tabla 1.

Ante la ausencia de estudios que demuestren la fiabilidad de la anterior clasificación, procedimos a medir la reproducibilidad y la consistencia interna del apartado radiológico de la clasificación de Coughlin y Shurnas.

Material y métodos

Seis traumatólogos (2 expertos, 2 sénior, 2 noveles) analizaron las radiografías anteroposterior y perfil en carga del pie de 70 pacientes afectos de HR y las catalogaron en base a la gradación de la clasificación de Coughlin y Shurnas en su apartado radiológico. Posteriormente y tras aleatorizar la muestra de radiografías, se repitió la observación pasados 3 días con tal de minimizar el sesgo de memoria en el retest. Con los datos obtenidos se analizó la fiabilidad inter-observador mediante el estadístico Fleiss Kappa (variables no dicotómicas y varios observadores) y el Coeficiente de Concordancia de Kendall usando el paquete estadístico



Figura 1 Ejemplo de ficha con radiografías anteroposterior y perfil que utilizaron los observadores para clasificar cada paciente.

Minitab 16.1.1. Dado que los grados 3 y 4 de la clasificación radiológica son iguales se establecieron 4 categorías a evaluar, el grado 0, 1, 2 y el 3 (fig. 1).

Tabla 1 Clasificación de Coughlin y Shurnas para la valoración de Hallux Rigidus

Grado	Arco de movilidad	Radiografía	Clínica
0	Flexión dorsal de 40-60° o pérdida de 10-20% comparando con el lado normal	Observaciones normales o mínimas	Ningún dolor subjetivo, solo rigidez; pérdida del movimiento pasivo en la exploración
1	Flexión dorsal de 30-40° o pérdida del 20-25% comparado con el lado normal	El espolón dorsal es el principal hallazgo, mínimo estrechamiento articular, esclerosis periarticular mínima, aplanamiento mínimo de la cabeza metatarsiana	Dolor leve o subjetivo ocasional y rigidez; dolor en los extremos de la flexión dorsal o en la flexión plantar en la exploración
2	Flexión dorsal de 10-30° o pérdida de 50-75% comparado con el lado normal	Los osteofitos dorsales, laterales y posiblemente mediales dan un aspecto aplanado a la cabeza del metatarsiano, no más del 25% de la región dorsal articular puede estar afectada en la zona lateral, los sesamoideos están aumentados de tamaño o son quísticos e irregulares	Dolor subjetivo casi constante y rigidez importante, dolor a lo largo de todo el arco de movilidad en la exploración (pero no en la zona media del arco)
3	Flexión dorsal de 10° o menos o pérdida del 75-100% comparado con el lado normal y pérdida notable de la flexión plantar (a menudo 10° o menos de flexión plantar)	Como en grado 2 pero estrechamiento sustancial, posiblemente cambios quísticos periarticulares, más del 25% de la región dorsal articular puede estar afectada en la zona lateral, los sesamoideos están aumentados de tamaño o son quísticos o irregulares	Dolor subjetivo casi constante y rigidez importante; dolor a lo largo de todo el arco de movilidad en la exploración (pero no en la zona media del arco)
4	Flexión dorsal de 10° o menos o pérdida del 75-100% comparado con el lado normal y pérdida notable de la flexión plantar (a menudo 10° o menos de flexión plantar)	Como en grado 2 pero estrechamiento sustancial, posiblemente cambios quísticos periarticulares, más del 25% de la región dorsal articular puede estar afectada en la zona lateral, los sesamoideos están aumentados de tamaño o son quísticos o irregulares	Dolor subjetivo casi constante y rigidez importante; dolor a lo largo de todo el arco de movilidad en la exploración y dolor claro en la zona media del arco

Resultados

En la tabla 2 y figura 2 podemos observar el porcentaje de acuerdo intraobservador que oscila entre el 68,57% (56,37-79,15) correspondiente al observador 6 y el 91,43% (82,27-96,79) del observador con más aciertos que fue el 3.

Sin embargo, estos resultados no son completamente ciertos al no tener en cuenta el análisis por porcentajes el factor azar. Para excluir dicho factor es necesario analizar los datos mediante el estadístico Fleiss Kappa obteniendo los siguientes resultados que muestra la tabla 3. El estadístico Kappa oscila en -1 y 1, correspondiendo el valor 0 al azar. Cuanto más próximo es el resultado al valor 1 mejor grado de concordancia.

El estadístico Fleiss Kappa no es capaz de discernir entre el grado de desacuerdo, es decir, no es lo mismo que dos observadores difieran puntuando una radiografía entre 1 y

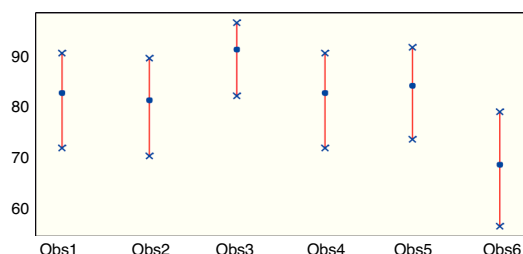


Figura 2 Gráfico en tanto por ciento de concordancia intraobservador en el text re-test (IC 95%).

2, que entre 1 y 3⁴. El grado de desacuerdo o la polaridad del desacuerdo para estos últimos es mucho mayor. Para ello usamos el Coeficiente de Concordancia de Kendall. Coeficiente, que tiene en cuenta la magnitud de la discordancia y que oscila entre valores de 0 y 1, indicando cuanto más cercano al 1 es el resultado, mayor grado de acuerdo⁵. Observamos los resultados en la tabla 4.

Procedemos a mostrar a continuación el mismo análisis de los datos pero en este caso para valorar la fiabilidad interobservadores (tablas 5 y 6).

Discusión

Ateniéndonos a los datos obtenidos y a la interpretación del estadístico Kappa propuesta por Landis y Koch⁶, así como los resultados del Coeficiente de Correlación de Kendall, podemos decir que el grado de concordancia interobservador de la clasificación de Coughlin y Shurnas es moderado (Kappa 0,46; Kendall 0,81), mientras que el comportamiento intraobservador es mucho mejor siendo este bueno (Kappa 0,52-0,84; Kendall 0,92-0,96) (tabla 7).

Quisimos valorar si la veteranía de los observadores podría influir en la concordancia interobservador, por lo que se procedió al análisis de los datos eliminando de los mismos a los 2 observadores noveles. El resultado del Fleiss Kappa fue de 0,49 y el Coeficiente Kendall de 0,83, con lo que podemos concluir que la influencia de la veteranía del observador en la concordancia interobservador es baja.

El comportamiento de la clasificación de Coughlin y Shurnas en su apartado radiológico es más que aceptable si observamos en la literatura estudios similares a este en el que

Tabla 2 Grado concordancia intraobservador en el text re-test

Observador	Casos analizados	Casos concordantes	Porcentaje	IC 95%
Obs1	70	58	82,86	(71,97, 90,82)
Obs2	70	57	81,43	(70,34, 89,72)
Obs3	70	64	91,43	(82,27, 96,79)
Obs4	70	58	82,86	(71,97, 90,82)
Obs5	70	59	84,29	(73,62, 91,89)
Obs6	70	48	68,57	(56,37, 79,15)

Tabla 3 Valores Fleiss Kappa intraobservador

Observador	Fleiss Kappa
Obs1	0,75
Obs2	0,71
Obs3	0,84
Obs4	0,73
Obs5	0,71
Obs6	0,52

Tabla 4 Valores de Coeficiente Correlación de Kendall en fiabilidad intraobservador

Observador	Coef. Kendall
Obs1	0,936701
Obs2	0,933382
Obs3	0,967651
Obs4	0,934881
Obs5	0,930546
Obs6	0,926163

Tabla 5 Porcentaje de acierto interobservadores. Muestra el % de casos en el que los 6 observadores coincidieron en los resultados

Casos analizados	Casos concordantes	Porcentaje	IC 95%
70	17	24,29	(14,83, 36,01)

Tabla 6 Índice Fleiss Kappa y Coeficiente Concordancia de Kendall interobservador

Fleiss Kappa	Coef. Kendall
0,46	0,818370

Tabla 7 Coeficiente Fuerza de la concordancia Kappa según Landis y Koch

< 0,00	Pobre (Poor)
0,01-0,20	Leve (Slight)
0,21-0,40	Aceptable (Fair)
0,41-0,60	Moderada (Moderate)
0,61-0,80	Considerable (Substantial)
0,81-1,00	Casi perfecta (Almost perfect)

se evalúan la concordancia de otras clasificaciones para otras patologías. Así, por ejemplo, Sayed-Noor et al. llevan a cabo un estudio similar al nuestro, en el que buscan evaluar la fiabilidad inter/intraobservador de varios sistemas de clasificación basados en la observación de pruebas de imagen en fracturas de calcáneo⁷. Así para la clasificación de Sanders obtienen un valor Kappa interobservador de 0,25 e intraobservador de 0,39 y para la clasificación de Letournel de 0,50 y 0,42 respectivamente: valores mucho menores a los valores Kappa de la clasificación Coughlin y Shurnas de nuestro estudio.

No hemos encontrado ningún estudio en la literatura que evalúe la fiabilidad de la clasificación Coughlin o de otras

clasificaciones para Hallux Rigidus por lo que no podemos establecer comparaciones.

Somos conscientes que la clasificación de Coughlin y Shurnas es una clasificación clínico-radiológica y que nuestro estudio evalúa únicamente la parte radiológica de la misma, por lo que los resultados deben aplicarse y tener en cuenta que corresponden a dicho apartado, y no extrapolarse a la clasificación al completo. Se precisaría de un estudio prospectivo para evaluar la parte clínica, y por tanto la clasificación al completo.

Conclusiones

A la luz de lo expuesto, podemos concluir que el comportamiento en cuanto a fiabilidad intraobservador del apartado radiológico de la clasificación de Coughlin y Shurnas es bueno mientras que el interobservador es moderadamente aceptable. Se necesitan más estudios que evalúen la clasificación al completo incluyendo su dimensión clínica, así como estudios de validez y fiabilidad para el resto de las clasificaciones para Hallux Rigidus existentes.

Bibliografía

1. Coughlin MJ, Shurnas PS. Hallux rigidus: demographics, etiology, and radiographic assessment. *Foot Ankle Int.* 2003;24(10):731-43.
2. Coughlin MJ, Shurnas PS. Hallux rigidus. Grading and long-term results of operative treatment. *J Bone Joint Surg Am.* 2003;85-A(11):2072-88.
3. Beeson P, Phillips C, Corr S, Ribbans W. Classification systems for hallux rigidus: a review of the literature. *Foot Ankle Int.* 2008;29(4):407-14.
4. Viera AJ, Garrett JM, Garret JM. Understanding interobserver agreement: the kappa statistic. *Fam Med.* 2005;37(5):360-3.
5. Keith M, Bower MS. Measurement system analysis with attribute data [consultado 14 Abr 2014]. Disponible en: http://www.minitab.com/uploadedFiles/Content/News/Published_Articles/measurement_system_analysis_attribute_data.pdf
6. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-74.
7. Sayed-Noor AS, Agren PH, Wretenberg P. Interobserver reliability and intraobserver reproducibility of three radiological classification systems for intra-articular calcaneal fractures. *Foot Ankle Int.* 2011;32(9):861-6.