

TÍTULO

¿El nuevo Sistema Unificado de Clasificación (SUC) de fracturas periprotésicas de cadera es más reproducible que la clasificación de Vancouver?

AUTORES

Rodrigo Jaramillo MD*, María Bautista MD MSc*, Jorge Rojas MD**, Meilyn Muskus MD**, Guillermo Bonilla MD***, Adolfo Llinás MD***, en nombre del Grupo de Investigadores ArtroRecon.

AFILIACIONES

* Departamento de Ortopedia y Traumatología, Hospital Universitario Fundación Santa Fe de Bogotá. Facultad de Medicina, Universidad del Rosario. Bogotá D.C., Colombia.

**Departamento de Ortopedia y Traumatología, Hospital Universitario Fundación Santa Fe de Bogotá. Bogotá D.C., Colombia.

*** Departamento de Ortopedia y Traumatología, Hospital Universitario Fundación Santa Fe de Bogotá. Facultad de Medicina, Universidad de Los Andes. Facultad de Medicina, Universidad del Rosario. Bogotá D.C., Colombia.

AUTOR DE CORRESPONDENCIA

Guillermo Bonilla MD. Departamento de Ortopedia y Traumatología, Hospital Universitario Fundación Santa Fe de Bogotá. Teléfono: 6030303 – ext. 5832. Dirección: Carrera 7 No. 117 - 15. Código postal: 110111186. Bogotá D.C., Colombia. Correo electrónico: bonillaguillermo@yahoo.com

RESUMEN

Introducción: Las fracturas periprotésicas son una complicación catastrófica que impacta la salud de los pacientes con artroplastias de cadera. El Sistema Unificado de Clasificación (SUC) fue desarrollado con el propósito de aplicar los principios de manejo de fracturas periprotésicas a cualquier reemplazo articular e incluir trazos de fracturas que el Sistema de Clasificación Vancouver (SCV) no contempla. El propósito de este estudio es evaluar y comparar la reproducibilidad de ambas clasificaciones.

Materiales y métodos: Diecinueve observadores (11 subespecialistas y 8 residentes) evaluaron las radiografías anteroposteriores de pelvis de 20 fracturas periprotésicas de cadera y clasificaron cada caso de acuerdo al SUC y al SCV. El acuerdo intraobservador e interobservador fue calculado mediante el estadístico AC2 de Gwet con ponderaciones para los acuerdos parciales.

Resultados: El promedio del acuerdo intraobservador de SUC fue 0.571 y 0.690 para SCV. El acuerdo interobservador para SUC fue 0.582 (IC 95 % 0.436 - 0.728) entre subespecialistas y 0.585 (IC 95 % 0.441 - 0.729) entre residentes. El acuerdo interobservador entre especialistas para SCV fue 0.637 (IC 95% 0.358 - 0.915) y 0.55 (IC 95% 0.307 - 0.795) entre residentes.

Discusión: La clasificación de SCU como muestra ser similar al SCV en términos de reproducibilidad. Existe una correlación entre el entrenamiento y una mayor reproducibilidad del uso de ambos sistemas. A pesar de la buena aceptación por parte de los observadores de este nuevo sistema de clasificación, es posible que su implementación en la práctica clínica requiera una mayor familiaridad con la misma.

Nivel de evidencia: III

PALABRAS CLAVE

Arthroplasty, Replacement, Hip; Reoperation; Periprosthetic Fractures; Classification;
Observer Variation.

INTRODUCCIÓN

La cirugía de reemplazo articular es uno de los procedimientos quirúrgicos más frecuentes en ortopedia (1,2). Se estima que para el año 2030 habrá un aumento del 174% en la demanda de artroplastias de cadera (3). En consecuencia, es probable que a medida que la población envejece se incremente la frecuencia de fracturas periprotésicas.

Las fracturas periprotésicas son aquellas que se producen alrededor de un implante o prótesis, o cuyo mecanismo de lesión está condicionado por la presencia de un implante (1). A pesar de que el reemplazo total de cadera es un procedimiento exitoso y tiene una baja incidencia de complicaciones, una fractura alrededor de un implante es una complicación catastrófica que impacta la salud de los pacientes (2).

Se estima que alrededor del 3.5% de las artroplastias de cadera experimentan una fractura periprotésica en los primeros 10 años posteriores a la cirugía (2) y se ha reportado una mortalidad asociada de hasta el 17% a los 6 meses y del 30% a 1 año (4). De acuerdo al reporte del 2011 del Registro Sueco de Cadera, las fracturas periprotésicas son la tercera causa de cirugía de revisión (5), mientras que en los registros del Reino Unido y Australia las fracturas periprotésicas representaron el 8.6% (6) y el 18.2% (7) de las revisiones respectivamente.

En la mayoría de los casos, este tipo de fracturas requieren de un manejo quirúrgico debido a las características del trazo de fractura y al compromiso de la estabilidad del implante (8). Así mismo, el tratamiento de este tipo de fracturas es un reto técnico para el cirujano, principalmente por la localización del trazo de fractura, la edad avanzada de los pacientes, la osteoporosis y las comorbilidades asociadas (1), por lo cual, los sistemas de clasificación tienen el papel fundamental de orientar este tratamiento (9).

Con el fin de ampliar y actualizar la clasificación de Vancouver (SCV) (Tabla 1) y aplicar los principios de tratamiento para todas las fracturas periprotésicas sin importar el hueso o articulación comprometida (10), en el 2014 Duncan (el mismo autor de la Clasificación de Vancouver) introdujo el Sistema Unificado de Clasificación de fracturas periprotésicas (SUC) (Tabla 2). Este nuevo sistema se diferencia del de Vancouver por incluir tres patrones de fractura adicionales. Además, busca estandarizar los principios de manejo de todas las fracturas periprotésicas e introduce un lenguaje sencillo y común para ayudar a la evaluación y el tratamiento de estas lesiones y, posiblemente, ayudar a la medición de los resultados (11).

Tanto el SUC como el SCV proponen algoritmos de tratamiento para las fracturas periprotésicas de acuerdo a cada categoría, a partir de evidencia de alto nivel (11,12). Sin embargo, hasta el momento no existen reportes en la literatura que validen la capacidad de cada escala para predecir dicho tratamiento; sobretodo en los trazos de fractura incluidos en la nueva escala (SUC).

El objetivo de este estudio es evaluar la validez, utilidad en la selección del tratamiento y percepción del Sistema Unificado de Clasificación y de la clasificación de Vancouver para fracturas periprotésicas de pelvis y fémur en reemplazo total de cadera en un grupo de cirujanos de cadera de diversos niveles de experiencia.

MATERIALES Y MÉTODOS

Casos y evaluadores

Para este estudio se evaluó el acuerdo interobservador de un sistema de clasificación. Se seleccionaron las radiografías anteroposterior (AP) de pelvis y AP y lateral de fémur de fracturas periprotésicas en reemplazos totales de cadera, incluyendo todo el espectro de severidad de las clasificaciones SCU y SCV, incorporando al menos 1 caso para cada categoría de clasificación. Los casos fueron seleccionados de la base de datos del registro del centro de cuidado clínico de reemplazos articulares de la Fundación Santa Fe de Bogotá. Tres autores (R.J., G.B., M.M.) seleccionaron y evaluaron las radiografías, verificando su adecuada calidad técnica para la inclusión. Mediante un consenso formal y de acuerdo con los hallazgos intraoperatorios se definió a que categoría del SCU o SCV pertenecía cada uno de los casos, considerando este como el estándar de oro para comparar el grado de precisión de la clasificación de los evaluadores.

Los observadores fueron seleccionados buscando tener el espectro completo de usuarios finales, incluyendo 11 subespecialistas con entrenamiento especial en cirugía de artroplastias con diversos grados de experiencia y 8 residentes de ortopedia en diferentes niveles de entrenamiento, quienes fueron invitados y aceptaron mediante consentimiento informado su participación en esta investigación. Ninguno de los evaluadores tuvo acceso a la identificación, nombres ni historia de los pacientes. Tanto los evaluadores como los casos recibieron números de codificación para ser identificados.

Radiografías y sesiones de evaluación

Todas las radiografías se encontraban en formato digital y se verificó que cumplieran con una adecuada calidad técnica que permitiera evaluar en su totalidad el segmento anatómico en cuestión. Para el fémur se presentaron radiografías en proyecciones AP y lateral, para la pelvis se presentaron radiografías en proyección AP.

La primera sesión de evaluación se realizó tanto para los subespecialistas como para los residentes en una reunión colectiva, realizando previo a la evaluación un entrenamiento y capacitación en los sistemas de clasificación, los cuales tuvieron disponibles durante todas las sesiones de evaluación; la evaluación se realizó de manera individual e independiente. Ocho semanas después de la primera evaluación, se presentaron los mismos casos a los observadores, alterando de manera aleatoria el orden de presentación para la evaluación a través del sistema de recolección electrónico Survey Monkey.

En la primera evaluación los observadores respondieron un cuestionario acerca de sus años de experiencia y de la percepción global subjetiva de los sistemas de clasificación.

Acuerdo intraobservador e interobservador

La afirmación común que Kappa es una “medida de acuerdo corregida de azar” es incorrecta. Como un estadístico de prueba, Kappa puede verificar que el acuerdo excede los niveles de azar; pero como una medida del nivel de acuerdo, kappa no es “corregido de azar”; por tanto, en ausencia de algún modelo explícito del proceso de toma de decisión de los evaluadores, de ninguna manera es claro como el azar afecta las decisiones de los participantes y como se podría corregir.

Kappa está influenciado por la prevalencia de la condición (distribución) y las tasas basales. Como un resultado, los kappas son raramente comparables entre estudios, procedimientos o

poblaciones. Kappa puede ser bajo aun cuando haya niveles altos de acuerdo y aun cuando las evaluaciones individuales sean muy precisas.

Hay una amplia controversia acerca de la utilidad del estadístico kappa para evaluar el acuerdo entre observadores y no debería ser visto como el estándar inequívoco o la manera predeterminada de cuantificar el acuerdo. Se debe usar un estadístico que es fuente de tanta controversia y se aconseja buscar alternativas.

Teniendo en cuenta lo anterior, se decidió basados en la naturaleza ordinal de los sistemas de clasificación utilizados y en el número de evaluadores, utilizar el estadístico AC2 propuesto por Gwet (2008), el cual es más robusto y menos propenso a paradojas que el Kappa.

Análisis Estadístico.

Se realizó un análisis del acuerdo intraobservador para cada grupo (subespecialistas y residentes) usando los valores del estadístico *AC2 Gwet* con sus correspondientes intervalos de confianza al 95% (IC 95%), con ponderaciones diseñadas específicamente teniendo en cuenta los acuerdos parciales de acuerdo a las similitudes fisiopatológicas e imagenológicas entre las categorías de las clasificaciones (Tabla 1 y Tabla 2).

Se evaluó el acuerdo intraobservador para todos los evaluadores, comparando los resultados del mismo evaluador en las dos evaluaciones realizadas. Para determinar el acuerdo intraobservador, se calculó el estadístico *AC2 Gwet* con sus correspondientes intervalos de confianza al 95%. Para todos los análisis estadísticos, se usó el software estadístico R (R Core Team (2013). R: A language and environment for statistical computing. R Foundation

for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>).

La interpretación del grado de acuerdo entre los observadores basado en la magnitud del coeficiente de acuerdo AC2 Gwet se realizó mediante el proceso de referenciación probabilístico recomendado por Gwet usando los niveles de acuerdo descritos por Altman (Tabla 3). La comparación directa clásica de los valores de los coeficientes de acuerdo con los puntos de referencia descritos, por ejemplo, los de Landis y Koch, da una caracterización excesivamente optimista del grado de acuerdo entre los evaluadores, ignorando los efectos adversos que pueden tener muestras pequeñas de evaluadores y/o casos en la precisión de los coeficientes de acuerdo.

RESULTADOS

Sistema de Clasificación Unificado (SUC)

Acuerdo interobservador

El acuerdo interobservador calculado estuvo entre un acuerdo discreto: 0.403 (IC 95%: 0.257 – 0.550) a un acuerdo moderado: 0.599 (IC 95%: 0.459 – 0.739) para los dos grupos de observadores (subespecialistas y residentes) en las dos evaluaciones realizadas. Los datos completos para cada grupo de evaluadores en cada evaluación, con las probabilidades acumuladas de pertenecer a los rangos de Altman se presentan en las Tabla 4 y 5.

No se observaron diferencias en el grado de acuerdo interobservador entre subespecialistas y residentes. En el grupo de subespecialistas se mantuvo el grado de acuerdo entre la primera y la segunda evaluación, sin embargo, entre los residentes el grado de acuerdo bajó en la segunda evaluación.

Acuerdo intraobservador

El promedio de acuerdo intraobservador calculado para el total de los evaluadores fue de 0.571 para un grado de acuerdo moderado. Hubo mayor grado de acuerdo intraobservador entre los subespecialistas que entre los residentes. En la Tabla 6 se presentan los datos para cada evaluador y los promedios por subgrupos.

Sistema de Clasificación de Vancouver (SCV)

Acuerdo interobservador

El acuerdo interobservador calculado estuvo entre un acuerdo discreto (0,492 IC 95% 0,304 - 0,679) a un acuerdo moderado (0,637 IC 95% 0,358 - 0,915) para los dos grupos de evaluadores (subespecialistas y residentes) en las dos evaluaciones realizadas. Los datos

completos para cada grupo de evaluadores en cada evaluación y con las probabilidades acumuladas de pertenecer a los rangos de Altman se presentan en las Tabla 7 y 8.

Al igual que en el SCU no se observaron diferencias en el grado de acuerdo interobservador entre subespecialistas y residentes. Entre la primera y la segunda evaluación no hubo una tendencia en el grado de acuerdo, mientras que para los subespecialistas bajó para los residentes subió.

Acuerdo intraobservador

El promedio de acuerdo intraobservador calculado para el total de los evaluadores fue de 0.69 para un grado de acuerdo sustancial. El grado de acuerdo fue similar entre residentes y subespecialistas. En la Tabla 9 se presentan los datos para cada evaluador y los promedios por subgrupos.

Correlación entre entrenamiento y acuerdo intraobservador.

Una prueba de correlación de Spearman se realizó para evaluar la relación entre los años de experiencia y el grado de reproducibilidad intraobservador para los dos sistemas de clasificación usando los 19 participantes (residentes y subespecialistas). No hubo correlación entre los años de experiencia y el grado de reproducibilidad intraobservador, $\rho = 0.1205$ $p = 0.623$.

DISCUSIÓN

Las fracturas periprotésicas parecen haber sido inicialmente reportadas en la literatura por Horwitz y Lenobel en 1954 (2). Actualmente, las fracturas periprotésicas son una causa importante de cirugía de revisión (13). Se estima que alrededor del 3.5% de los pacientes con artroplastias primarias experimentan una fractura periprotésica en los primeros 10 años posterior a la cirugía, de las cuales el 80% requieren manejo operatorio (14). Adicionalmente, la tasa acumulativa de mortalidad es del 21%, con una mortalidad a 1 año de 11% (15,16).

Esto significa que las fracturas periprotésicas de cadera representan una complejidad adicional que requieren toda la atención de los especialistas en reemplazos articulares y deben ser reconocidas de forma adecuada por los cirujanos ortopedistas en entrenamiento.

Los principios del tratamiento de las fracturas periprotésicas de cadera se basan en la localización del trazo de fractura, la estabilidad del implante y la reserva ósea del paciente (10,12,17). Por lo tanto, el objetivo de cualquier sistema de clasificación de fracturas periprotésicas es permitir la identificación de cada caso dentro de una categoría que permita guiar su tratamiento, establecer el pronóstico y la facilitar la comunicación entre cirujanos (9).

El SCV, desarrollado por Duncan y Masri, está basado en 3 aspectos: localización, estabilidad del implante y la masa ósea circundante y a través de los años ha sido este el sistema más universalmente aceptado y el más utilizado (18). Adicionalmente, presentaba una reproducibilidad inter-observador (κ : 0.69) e intra-observador (κ : 0.61) aceptables (11,19). Sin embargo, en el 2014, el mismo autor de esta clasificación introdujo el SUC, con el fin de ampliar y actualizar el SCV y aplicar los principios de tratamiento

para todas las fracturas periprotésicas, incluyendo otras articulaciones diferentes a cadera (10) e introducir un lenguaje sencillo y común para ayudar a la evaluación y el tratamiento de estas lesiones y a la medición de sus resultados (11).

A pesar de la aparente similitud de los dos sistemas de clasificación, la inclusión de tres patrones adicionales de fractura, los tipos D, E y F, que tienen principios únicos en su tratamiento (10), para un total de 14 categorías, puede suponer un retro para la reproducibilidad del SUC en comparación con el SCV.

En reportes previos en la literatura, el SUC demostró tener acuerdos inter- e intra-observador “casi perfectos”, sin diferencias estadísticamente significativas entre expertos y no expertos (11), posiblemente debido a que este estudio fue realizado por los desarrolladores de la clasificación (20). Por el contrario, en nuestro estudio se obtuvieron acuerdos entre moderados y sustanciales y se encontró una correlación con la presencia o ausencia de entrenamiento en subespecialidad. Esto indica que la escala puede ser difícil usar y que su implementación en la práctica clínica requiere un estudio más intenso. Es decir, su reproducibilidad está estrechamente relacionada con la familiaridad de los cirujanos con las categorías de la nueva clasificación. Lo que también se puede concluir a partir del hecho de que en este estudio, los acuerdos inter-observador e intra-observador tanto en subespecialistas como en residentes fueron ligeramente superiores para SCV.

Las clasificaciones de SCV y SUC evaluadas en un grupo de 19 evaluadores con diversos grados de entrenamiento muestran una reproducibilidad discreta a moderada. Se requieren estrategias para mejorar el grado de reproducibilidad de estos sistemas de clasificación. No hay diferencias en el grado de acuerdo interobservador entre las clasificaciones SUC y SCV. El SCV muestra un mayor grado de acuerdo intraobservador (sustancial) con respecto

a la clasificación SUC (moderado). No hubo correlación entre los años de experiencia y el grado de reproducibilidad intraobservador.

En conclusión, la clasificación de SCU como nuevo sistema de clasificación muestra ser similar al SCV en términos de reproducibilidad y existe una correlación entre el entrenamiento y una mayor reproducibilidad del uso de ambos sistemas. A pesar de la buena aceptación por parte de los observadores de este nuevo sistema de clasificación, es posible que su implementación en la práctica clínica requiera una mayor familiaridad con la misma.

REFERENCIAS

1. Della Rocca GJ, Leung KS, Pape H-C. Periprosthetic fractures: epidemiology and future projections. *J Orthop Trauma*. 2011 Jun;25 Suppl 2(6):S66-70.
2. Lindahl H. Epidemiology of periprosthetic femur fracture around a total hip arthroplasty. *Injury*. 2007 Jun;38(6):651-4.
3. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Jt Surg*. 2007;89(4):780-5.
4. Sarmah SS, Patel S, Reading G, El-Husseiny M, Douglas S, Haddad FS. Periprosthetic fractures around total knee arthroplasty. *Ann R Coll Surg Engl*. 2012 Jul;94(5):302-7.
5. Garellick G, Kärrholm J, Lindahl H, Malchau H, Rogmark C, Rolfson O. Swedish Hip Arthroplasty Register: Annual Report 2013. 2013.
6. National Joint Registry for England, Wales, Northern Ireland and the Isle of Man. 12th Annual Report 2015. 2015.
7. Australian Orthopaedic Association. National Joint Replacement Registry. Hip and Knee Arthroplasty: Annual Report 2015. 2015.
8. Lindahl H, Garellick G, Regnér H, Herberts P, Malchau H. Three hundred and twenty-one periprosthetic femoral fractures. *J Bone Joint Surg Am*. 2006 Jun;88(6):1215-22.
9. Fletcher RH, Fletcher SW, Fletcher GS. *Clinical Epidemiology: The Essentials* 5th Edition. 2014. 1-251 p.
10. Huang JF, Shen JJ, Chen JJ, Zheng Y, Du WX, Liu FC, et al. New fracture pattern focusing on implant fracture for periprosthetic femoral fractures. *Int Orthop*. 2015;39(9):1765-9.
11. Vioreanu MH, Parry MC, Haddad FS, Duncan CP. Field testing the Unified Classification System for peri-prosthetic fractures of the pelvis and femur around a total hip replacement: An international collaboration. *Bone Jt J*. 2014;96B(11):1472-

- 7.
12. Difazio FA, Incavo SJ. Periprosthetic fracture after total hip arthroplasty. *Semin Arthroplasty*. 2005;16(2):119–26.
13. Duncan CP, Haddad FS. The Unified Classification System (UCS): improving our understanding of periprosthetic fractures. *Bone Joint J*. 2014 Jun;96–B(6):713–6.
14. Katz JN, Wright E a, Polaris JJZ, Harris MB, Losina E. Prevalence and risk factors for periprosthetic fracture in older recipients of total hip replacement: a cohort study. *BMC Musculoskelet Disord*. 2014 Jan;15(1):168.
15. Bhattacharyya T, Chang D, Meigs JB, Estok DM, Malchau H. Mortality after periprosthetic fracture of the femur. *J Bone Joint Surg Am*. 2007 Dec;89(12):2658–62.
16. Chimutengwende-Gordon M, Khan W, Johnstone D. Recent advances and developments in knee surgery: principles of periprosthetic knee fracture management. *Open Orthop J*. 2012 Jan;6:301–4.
17. Durante R, Hip DO. Fracturas periprotésicas de cadera. 2014;28(2):77–81.
18. Gaski GE, Scully SP. In brief: classifications in brief: Vancouver classification of postoperative periprosthetic femur fractures. *Clin Orthop Relat Res*. 2011 May;469(5):1507–10.
19. Naqvi GA, Baig SA, Awan N. Interobserver and Intraobserver Reliability and Validity of the Vancouver Classification System of Periprosthetic Femoral Fractures After Hip Arthroplasty. *J Arthroplasty*. Elsevier Inc.; 2012;27(6):1047–50.
20. Campbell DG, Garbuz DS, Masri BA, Duncan CP. Reliability of acetabular bone defect classification systems in revision total hip arthroplasty. *J Arthroplasty*. 2001;16(1):83–6.

Tabla 2. Matriz de ponderaciones para acuerdos parciales del SCV.

	Vancouver A	Vancouver B1	Vancouver B2	Vancouver B3	Vancouver C
Vancouver A	1	0	0	0	0
Vancouver B1	0	1	0.5	0.25	0
Vancouver B2	0	0.5	1	0.25	0
Vancouver B3	0	0.25	0.5	1	0
Vancouver C	0	0	0	0	1

Tabla 3. Interpretación de rangos de acuerdo según Altman

Valor Coeficiente	Interpretación
< 0.20	Acuerdo insignificante
0.21 – 0.40	Acuerdo discreto
0.41 – 0.60	Acuerdo moderado
0.61 – 0.80	Acuerdo sustancial
0.81 – 1.00	Acuerdo casi perfecto

Tabla 4. Resultados del acuerdo interobservador

Grupo Observador	Evaluación	AC2 Gwet	IC 95%	Grado de acuerdo
Subespecialistas	1	0.582	0.436 – 0.728	Moderado
	2	0.599	0.459 – 0.739	Moderado
Residentes	1	0.585	0.441 – 0.729	Moderado
	2	0.403	0.257 – 0.550	Discreto

Tabla 5. Probabilidades acumuladas de pertenecer a los rangos de referencia de Altman

Grupo observador	Evaluación	Rango de referencia	Descripción	Ac2
Subespecialistas	1	0.81 A 1	Casi perfecto	0,000569499
		0.61 A 0.8	Sustancial	0,394096931
		0.41 A 0.6	Moderado	0,996700372
		0.21 A 0.4	Discreto	0,999999994
		<0.2	Insignificante	1
	2	0.81 A 1	Casi perfecto	0,000842939
		0.61 A 0.8	Sustancial	0,49376678
		0.41 A 0.6	Moderado	0,999062582
		0.21 A 0.4	Discreto	1
		<0.2	Insignificante	1
Residentes	1	0.81 A 1	Casi perfecto	0,000666091
		0.61 A 0.8	Sustancial	0,411425107
		0.41 A 0.6	Moderado	0,997120477
		0.21 A 0.4	Discreto	0,999999995
		<0.2	Insignificante	1
	2	0.81 A 1	Casi perfecto	1,55795E-09
		0.61 A 0.8	Sustancial	0,001639481
		0.41 A 0.6	Moderado	0,51785712
		0.21 A 0.4	Discreto	0,998776627
		<0.2	Insignificante	0,999999999

Tabla 6. Acuerdo intraobservador para cada evaluador, promedios por subgrupos y promedio total.

Grupo Observador	No. observador	AC2 Gwet	Promedio por grupo	Promedio global
Subespecialistas	1	0.529	0.627	0.571
	2	0.607		
	3	0.617		
	4	0.794		
	5	0.626		
	6	0.38		
	7	0.755		
	8	0.656		
	9	0.793		
	10	0.685		
	11	0.46		
Residentes	12	0.529	0.49325	
	13	0.518		
	14	0.518		
	15	0.431		
	16	0.229		
	17	0.574		
	18	0.581		
	19	0.566		

Tabla 7. Resultados del acuerdo interobservador

Grupo Observador	Evaluación	AC2 Gwet	IC 95%	Grado de acuerdo
Subespecialistas	1	0.637	0.358 – 0.915	Moderado
	2	0.492	0.304 – 0.679	Discreto
Residentes	1	0.55	0.307 – 0.795	Discreto
	2	0.519	0.364 – 0.673	Moderado

Tabla 8. Probabilidades acumuladas de pertenecer a los rangos de referencia de Altman

Grupo observador	Evaluación	Rango de referencia	Descripción	Ac2
Subespecialistas	1	0.81 A 1	Casi perfecto	0.0757
		0.61 A 0.8	Sustancial	0.6265
		0.41 A 0.6	Moderado	0.9805
		0.21 A 0.4	Discreto	0.9992
		<0.2	Insignificante	0.9993
	2	0.81 A 1	Casi perfecto	3.16712E-05
		0.61 A 0.8	Sustancial	0.080368463
		0.41 A 0.6	Moderado	0.883918425
		0.21 A 0.4	Discreto	0.999925343
		<0.2	Insignificante	1
Residentes	1	0.81 A 1	Casi perfecto	0.005778018
		0.61 A 0.8	Sustancial	0.306758941
		0.41 A 0.6	Moderado	0.93513024
		0.21 A 0.4	Discreto	0.999793644
		<0.2	Insignificante	0.999997245
	2	0.81 A 1	Casi perfecto	4.09192E-06
		0.61 A 0.8	Sustancial	0.099271397
		0.41 A 0.6	Moderado	0.970546641
		0.21 A 0.4	Discreto	0.999999794
		<0.2	Insignificante	1

Tabla 9. Acuerdo intraobservador para cada evaluador, promedios por subgrupos y promedio total.

Grupo Observador	No. observador	AC2 Gwet	Promedio por grupo	Promedio global
Subespecialistas	1	0.588	0.685	0.690
	2	0.802		
	3	0.761		
	4	0.677		
	5	0.68		
	6	0.225		
	7	1		
	8	0.079		
	9	1		
	10	1		
	11	0.723		
Residentes	12	0.901	0.697375	0.690
	13	0.878		
	14	0.683		
	15	0.226		
	16	0.588		
	17	0.685		
	18	0.757		
	19	0.861		