



## ORIGINAL BREVE

## Fiabilidad interobservador de los 4 índices de comorbilidad más utilizados en pacientes ancianos

Marco A. Zelada Rodríguez<sup>a,\*</sup>, Javier Gómez-Pavón<sup>b</sup>, Pilar Sorando Fernández<sup>b</sup>,  
Andrés Franco Salinas<sup>b</sup>, Laurennny Mercedes Guzmán<sup>b</sup> y Juan José Baztán<sup>b</sup>

<sup>a</sup> Servicio de Geriatria, Hospital de Sant Llàtzer, CST, Terrassa, Barcelona, España

<sup>b</sup> Servicio de Geriatria, Hospital Central de la Cruz Roja, Madrid, España

## INFORMACIÓN DEL ARTÍCULO

## Historia del artículo:

Recibido el 27 de junio de 2011

Aceptado el 26 de septiembre de 2011

On-line el 20 de enero de 2012

## Palabras clave:

Comorbilidad

Fiabilidad interobservador

Ancianos

Cuidados agudos

## RESUMEN

**Objetivo:** Conocer la fiabilidad interobservador de los 4 índices de comorbilidad más utilizados en ancianos: índice de Charlson (ICH), *Cumulative Illness Rating Scale for Geriatrics* (CIRS-G), índice de Kaplan-Feinstein (IKF), e índice de coexistencia de enfermedad (ICED).

**Material y métodos:** Cuatro médicos, previamente entrenados, revisaron de forma independiente 40 historias clínicas de pacientes mayores de 75 años, ingresados por patología médica aguda, realizando los 4 índices y cronometrando el tiempo. Se analizó el coeficiente de correlación intraclass (CCI) para los índices cuantitativos (ICH y CIRS-G) y el coeficiente Kappa para índices cualitativos (IKF e ICED), las concordancias <0,4 se consideraron deficientes; 0,4-0,75 aceptable, y >0,75 excelente.

**Resultados:** Los pacientes de las historias evaluadas tenían una edad media de 85,93 ( $\pm 5,35$ ) años, siendo el 72,5% mujeres. El CCI global de los 4 evaluadores para el ICH fue 0,78 (IC del 95%: 0,67-0,86) y para el CIRS-G (score): 0,66 (IC del 95%: 0,53-0,78). Los valores del coeficiente Kappa para el IKF oscilaron entre 0,51-0,76 y para el ICED entre 0,44-0,66. El tiempo de aplicación fue menor para el ICH (mediana de 39 segundos [30-45]) e IKF (42 segundos [35-52]) y mayor para el CIRS-G (score) (128 segundos [110-160]) e ICED (102 segundos [80-124]).

**Conclusiones:** De los 4 índices valorados, el ICH y el índice CIRS-G (score), son los que presentan una mejor fiabilidad interobservador. El ICH y el IKF, presentan menor tiempo de aplicación.

© 2011 SEGG. Publicado por Elsevier España, S.L. Todos los derechos reservados.

## The interrater reliability of four common comorbidity indexes used in elderly patients

## ABSTRACT

## Keywords:

Comorbidity

Interrater reliability

Elderly

Acute care

**Objective:** To report on the interrater reliability of four common comorbidity indexes used in the hospitalised elderly: Charlson Index (CI), Geriatric Cumulative Illness Rating Scale (CIRS-G), Index of Co-existent Disease (CoD) and Kaplan-Feinstein Index (KFI).

**Method:** Four trained observers, independently reviewed the same 40 medical charts of hospitalised geriatric patients. Scores for the four indexes were calculated, along with the intraclass correlations coefficient (ICC) (quantitative index: CI and CIRS-G) and Kappa coefficient (qualitative index: CoD and KFI). The agreement <0.4 was considered deficient, 0.4-0.75 acceptable and >0.75 excellent.

**Results:** A total of 40 patients (29 women) of 85.93 ( $\pm 5.35$ ) years were analysed. Intraclass correlations coefficient: CI: 0.78 (95% CI: 0.67-0.86); CIRS-G (score): 0.66 (95% CI: 0.53-0.78). Kappa coefficient: KFI: 0.51 to 0.76; CoD: 0.44-0.66. The application time was lower for the Charlson index (median of 39 seconds [30-45]) and the KFI (42 seconds [35-52]) and higher for CIRS-G (score) (128 seconds [110-160]) and CoD (102 seconds [80-124]).

**Conclusions:** Of the four comorbidity indexes used in a hospitalised elderly population, the CI, and CIRS-G (score), are those that have better interrater reliability. The Charlson index and KFI show a lower application time than the CIRS-G (score).

© 2011 SEGG. Published by Elsevier España, S.L. All rights reserved.

\* Autor para correspondencia.

Correo electrónico: MAZelada@cst.cat (M.A. Zelada Rodríguez).

## Introducción

La comorbilidad es un problema frecuente que contribuye a la complejidad de los pacientes ancianos. Se ha descrito que el 24% de los mayores de 65 años y el 31,4% de los mayores de 85 padecen cuatro o más patologías crónicas<sup>1,2</sup>.

Se entiende por comorbilidad «cualquier entidad adicional que ha existido o puede ocurrir, durante el curso clínico de un paciente con una enfermedad guía»<sup>2,3</sup>. Su importancia radica en que puede retrasar e influir en el diagnóstico clínico, en las decisiones de tratamiento, asociarse a complicaciones, afectar la recuperación funcional, empeorar la calidad de vida, incrementar el uso de recursos sanitarios, alterar la supervivencia y actuar como variable moduladora relevante de resultados en los estudios clínicos en ancianos<sup>1-4</sup>.

Los índices de comorbilidad buscan reducir las enfermedades de una persona y su severidad a una puntuación que permita la comparación con otras personas. Ante la existencia de diferentes métodos para valorar la comorbilidad, revisiones previas han encontrado 12 índices válidos, de los cuales teniendo en cuenta una serie de criterios (validez concurrente, predictiva y de constructo, fiabilidad inter e intraobservador, estrategias analíticas, métodos de obtención de datos, frecuencia de estudios publicados, etc.), han seleccionado 4 índices como los más apropiados para la evaluación de la comorbilidad en pacientes ancianos: índice de Charlson (ICH), *Cumulative Illness Rating Scale-Geriatrics* (CIRS-G), *Index of Coexistent Disease* (ICED) y el índice de Kaplan-Feinstein (IKF)<sup>1-4</sup>.

La reproducibilidad de dichos índices no ha sido evaluada en nuestro medio y específicamente en pacientes ancianos mayores de 75 años. Dada su importancia clínica, decidimos realizar este estudio con el objetivo de conocer la fiabilidad interobservador de los 4 índices de comorbilidad más utilizados.

## Material y métodos

Inicialmente se obtuvieron los manuales originales en formato electrónico de los índices de ICH, CIRS-G, y el IKF, siendo el manual del ICED proporcionado directamente por los autores. Posteriormente, tanto los índices como sus manuales de aplicación fueron traducidos por dos investigadores y las discrepancias resueltas por consenso. Dado que los índices a evaluar consistían en una lista de patologías médicas, no se consideró necesario realizar una posterior retrotraducción.

Las características de los índices utilizados fueron:

- Índice de Charlson es el más conocido, evalúa 16 enfermedades que están explícitamente definidas y puntúa mediante una variable continua desde 0 a 31<sup>5</sup>.
- El CIRS-G contempla 14 grupos de enfermedades o categorías, evaluando su severidad de 0 a 4, dando una puntuación final (score) que es la suma de dichas puntuaciones (entre 0-56). Se considera baja comorbilidad valores «0-2», media: «3-8», alta > 8<sup>6</sup>. Otras formas de expresar su puntuación que no se han utilizado en el presente estudio (índice de severidad, el número de categorías de severidad 3 y 4).
- El ICED es el único que valora el estado funcional. Consta de dos secciones: 14 ítems para la sección de enfermedades, con una puntuación de severidad de 0 a 4, y 12 ítems para la sección de compromiso físico (estado funcional), con una puntuación de severidad de 0 a 2. La puntuación final agrupa las puntuaciones de ambas secciones y oscila entre 0-37<sup>7</sup>.
- El índice de Kaplan-Feinstein consta de 12 ítems por sistemas de enfermedades explícitamente definidas, con una severidad de 0 a 3 para cada sistema. La puntuación final es una variable ordinal de 4 categorías (de categoría 0 a categoría 3), en donde si

**Tabla 1**

Fiabilidad interobservador de los índices de comorbilidad y tiempo de los evaluadores en segundos empleado en la aplicación de los índices

	CCI (IC del 95%)	Tiempo (s) <sup>a</sup>
Índice de Charlson	0,78 (0,67-0,86)	39 s (30-45)
Índice de CIRS-G (score)	0,66 (0,53-0,78)	128 s (110-160)
Índice de Kaplan-Feinstein	Kappa 0,51-0,76	42 s (35-52)
Índice de coexistencia de enfermedad (ICED)	0,44-0,66	102 s (80-124)

CCI: coeficiente de correlación intraclase; CIRS-G (score): *Cumulative Illness Rating Scale for Geriatrics*; DE: desviación estándar; IC: intervalo de confianza; s: segundos.

<sup>a</sup> Expresado en mediana (rango intercuartílico percentil 25-75%).

están presentes múltiples enfermedades, la enfermedad con el más alto rango define la categoría, excepto si dos o más enfermedades están en grado 2 de severidad en donde la categoría final es 3<sup>8</sup>.

Para la formación de los 4 investigadores (un médico especialista en Geriatria y tres médicos residentes de Geriatria de tercer y cuarto año) se elaboró una guía resumen de cada índice y se discutió las puntuaciones en dos sesiones, con la aplicación posterior de dichos índices en 5 historias clínicas de pacientes ancianos ingresados por patología médica aguda.

Para este estudio se seleccionaron 40 historias de pacientes mayores de 75 años, ingresados consecutivamente en la unidad de agudos del servicio de Geriatria del Hospital Central de la Cruz Roja de Madrid, durante el mes de mayo de 2009; mediante un muestreo por conveniencia, el tamaño de la muestra se estableció en base al estudio realizado por Hall SF, et al.<sup>9</sup>. Estas historias fueron revisadas de forma independiente y simultánea por cada investigador extrayendo los datos necesarios para completar los ítems de cada índice de comorbilidad, cronometrando el tiempo utilizado en cada índice, el análisis estadístico lo realizó otro investigador independiente.

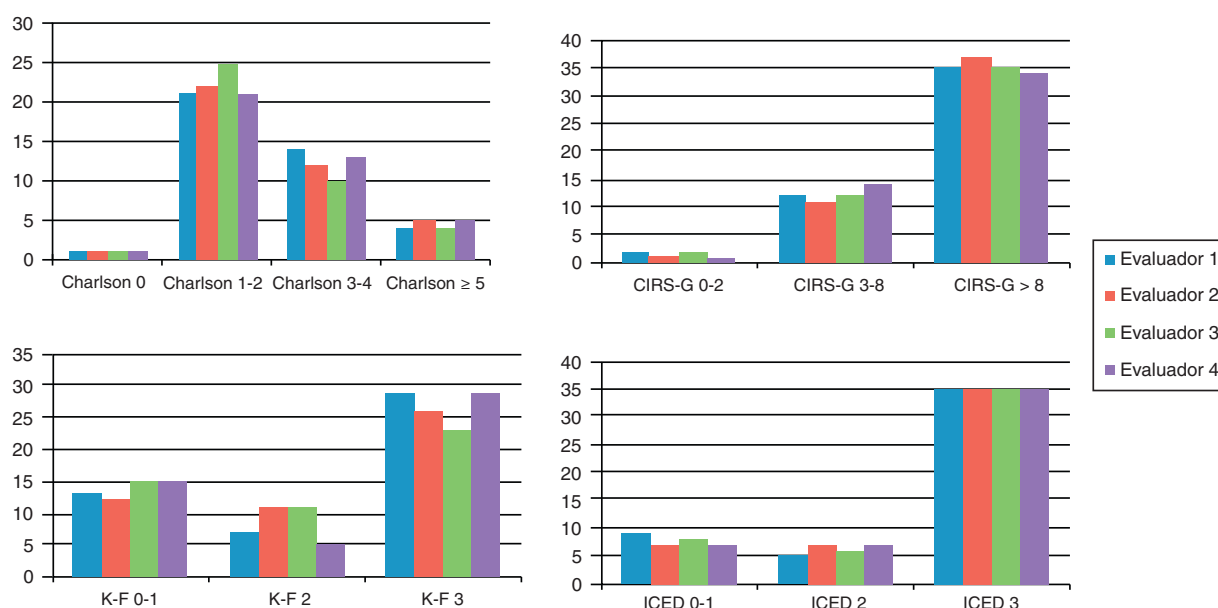
## Análisis estadístico

La fiabilidad interobservador se analizó con el coeficiente de correlación intraclase (CCI) para los índices cuantitativos (ICH y CIRS-G) y el coeficiente Kappa para índices cualitativos (IK-F e ICED). Siendo ambos coeficientes directamente comparables. Se consideró <0,4 como concordancia deficiente, 0,4-0,75 concordancia aceptable y >0,75 concordancia excelente<sup>10</sup>. El tiempo de aplicación fue medido en forma de mediana y rango intercuartílico. Para el análisis estadístico se utilizó el paquete estadístico SPSS versión 15.

## Resultados

De los 40 pacientes estudiados 29 (72,5%) eran mujeres, edad media de 85,93 ( $\pm 5,35$ ) años (rango de 75 a 98 años), con una situación funcional basal previa al ingreso medida por el índice de Barthel de  $61,81 \pm 36,34$  y mental medida por la escala de incapacidad psíquica de Cruz Roja de  $1,34 \pm 1,67$ . El principal motivo de ingreso fue infección (51%) seguido de insuficiencia cardíaca (22,4%).

El grado de concordancia entre observadores y el tiempo promedio de los evaluadores empleados en la aplicación de los índices se muestra en la **tabla 1**. El índice de Charlson muestra una concordancia excelente seguido del CIRS-G (score) con concordancia aceptable. En relación con el tiempo de aplicación de los 4 índices fue sensiblemente menor fue para el ICH y el K-F.



**Figura 1.** Distribución de la frecuencia de las categorías de los cuatro índices de comorbilidad según la puntuación de los diferentes evaluadores.

La distribución de puntuaciones de los diferentes evaluadores, para los 4 índices de comorbilidad estudiados, se presenta en la figura 1.

## Discusión

La muestra del presente estudio es una población eminentemente geriátrica, constituida por pacientes muy ancianos ingresados por patología médica aguda (rango de edad de 75–98 años), con presencia de dependencia funcional moderada y frecuente deterioro cognitivo.

Los niveles de fiabilidad obtenidos para los 4 índices de comorbilidad fueron excelente para el ICh y aceptable para CIRS-G (score), IKF e ICED, muy similares a los presentados por Hall SF, et al.<sup>9</sup>, pero ligeramente inferiores a los encontrados en otros estudios<sup>6,12–14</sup>. Debido a la falta de uniformidad en los reportes de los estudios de fiabilidad es difícil establecer comparaciones<sup>9,11</sup>; sin embargo, creemos que las diferencias encontradas podrían explicarse por 4 razones:

En primer lugar los pacientes de la muestra estudiada son sensiblemente más envejecidos (media de 86 años) que los del trabajo de Extermann M (media de 75 años)<sup>12</sup>, y otros estudios de edad media inferior<sup>6,13,14</sup>.

En segundo lugar, tenían mayor carga de enfermedad ya que puntuaban cero según el ICh solo el 2,4%, respecto al 50% en el estudio de Hall SF, et al.<sup>9</sup> y 64% en el de Extermann M<sup>12</sup>, lo cual aumenta la probabilidad de discrepancias entre los observadores, sobre todo para los índices CIRS-G, IKF e ICED, que incluyen variables funcionales, donde un buen nivel funcional mostraría menos errores entre los diferentes observadores.

En tercer lugar, los pacientes de nuestra muestra presentaban mayor deterioro funcional (índice de Barthel 61,81 ± 36,34). La mayoría de los estudios no presentan la situación funcional, encontramos en el estudio de Salvi F et al.<sup>13</sup> (índice de Barthel 87,1 ± 21,5). Esto afectaría sobre todo al CIRS-G, IKF e ICED, que incluyen variables funcionales, donde un buen nivel funcional mostraría menos errores entre los diferentes observadores.

Finalmente, a diferencia de los demás estudios que utilizaron informes de alta o bases de datos de pacientes ambulatorios<sup>6,9,12</sup> o ingresados<sup>6,13,14</sup>, en el presente estudio se utilizó la historia clínica completa la cual, al tener mayor cantidad de información y un perfil

más complejo, si bien contribuiría a la mayor riqueza de los datos sobre todo en el CIRS-G e ICED, podría ser un factor de variabilidad entre observadores.

Respecto a las diferencias entre las puntuaciones de los índices de comorbilidad estudiados, tal como se ilustra en la figura 1, el CIRS-G, al ser más específico de la población anciana, con afecciones geriátricas evaluadas según diferentes grados de severidad, hace puntuar a casi todos los sujetos, mientras que el ICh tiende a establecer puntuaciones menores, aunque menos sesgado a la izquierda que en otros estudios con pacientes más jóvenes<sup>6,12–14</sup>. De igual forma, con el ICED y el IKF que puntúan deterioro funcional, dada la alta prevalencia de discapacidad, aparece un sesgo hacia la derecha, es decir hacia su máxima puntuación (categoría de valor 3).

En relación con el tiempo de aplicación, no hemos encontrado estudios comparativos. Los promedios obtenidos muestran diferencias importantes entre uno y otro índice, ocupando la aplicación del ICh e IKF tres veces menos tiempo que el CIRS-G (score) e ICED.

En conclusión, de los 4 índices de comorbilidad más usados en pacientes muy ancianos hospitalizados por patología médica aguda, encontramos una excelente concordancia interobservador para el índice de Charlson y aceptable para el CIRS-G (score), ICED y Kaplan-Feinstein, siendo los índices de Charlson y de Kaplan-Feinstein los que requieren un menor tiempo de aplicación. Por lo que podrían ser utilizados, según los objetivos del clínico y/o investigador, siendo necesario realizar estudios comparativos para valorar la capacidad predictiva de dichos índices.

## Bibliografía

- Martínez Velilla NI, Gaminde Inda ID. Comorbidity and multimorbidity indexes in the elderly patient. *Med Clin (Barc)*. 2011;136:441–6.
- Abizanda P, Paterna GM, Martínez SE, López JE. Evaluación de la comorbilidad en la población anciana: utilidad y validez de los instrumentos de la medida. *Rev Esp Geriatr Gerontol*. 2010;45:219–28.
- De Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity: a critical review of available methods. *J Clin Epidemiol*. 2003;56:221–9.
- Lash TL, Mor V, Wieland D, Ferrucci L, Satariano W, Silliman RA. Comorbidity. Review article. Methodology, design and analytic techniques to address measurement of comorbid disease. *J Gerontol*. 2007;62A:281–5.
- Charlson ME, Pompei P, Ales KL, Mackenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40:373–83.

6. Miller MD, Paradis CF, Houck PR, Mazumdar S, Stack JA, Rifai AH, et al. Rating chronic medical illness burden in geropsychiatric patients: application of the Cumulative Illness Rating Scale. *Psychiatry Res.* 1992;41:237–48.
7. Cleary PD, Greenfield S, Mulley AG, Pauker SG, Schroeder SA, Wexler L, et al. Variations in length of stay and outcomes for six medical and surgical conditions in Massachusetts and California. *JAMA.* 1991;266:73–9.
8. Kaplan MH, Feinstein AR. The importance of classifying initial comorbidity in evaluating the outcome of diabetes mellitus. *J Chronic Dis.* 1974;27:387–404.
9. Hall SF, Groome P, Streiner DL, Rochon P. Interrater reliability of measurement of comorbid illness should be reported. *J Clin Epidemiol.* 2006;59:926–33.
10. Hernández Aguado I, Porta Serra M, Miralles M, García Benavides F, Bolívar F. La cuantificación de la variabilidad en las observaciones clínicas. *Med Clin.* 1990;95:424–9.
11. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64:96–106.
12. Extermann M. Measuring comorbidity in older cancer patients. *Eur J Cancer.* 2000;36:453–71. Review.
13. Salvi F, Miller MD, Grilli A, Giorgi R, Towers AL, Morichi V, et al. A Manual of guidelines to score the modified cumulative illness rating scale and its validation in acute hospitalized elderly patients. *J Am Geriatr Soc.* 2008;56:1926–31.
14. Waite K, Oddone E, Weinberger M, Samsa G. Lack of association between patients' measured burden of disease and risk for hospital readmission. *J Clin Epidemiol.* 1994;47:1229–36.