

¿Es el factor de impacto un índice apropiado para determinar el grado de evidencia de estudios sobre procedimientos terapéuticos en revistas quirúrgicas?

Carlos Manterola^{a,b}, Viviana Pineda^a, Manuel Vial^{a,b} y Héctor Losada^a

^aDepartamento de Cirugía. Universidad de la Frontera. Temuco. Chile.

^bCIGES (Capacitación, Investigación y Gestión para la Salud Basada en Evidencia). Universidad de La Frontera. Temuco. Chile.

Resumen

Introducción. El objetivo de este estudio es determinar la asociación entre el factor de impacto (FI) de las revistas quirúrgicas y el grado de evidencia de los artículos publicados en ellas sobre procedimientos terapéuticos.

Material y método. Estudio bibliométrico. Se incluyeron revistas quirúrgicas registradas en ISI® 2002, que fueron agrupadas según el percentil 33 de su FI en: grupo 1 (0,128 a 1,690), grupo 2 (1,691 a 2,886) y grupo 3 (2,887 a 6,674). Se aplicó un muestreo aleatorio simple de las revistas de cada grupo, con lo que los grupos quedaron conformados por: grupo 1: *European Journal of Surgery, American Surgeon y World Journal of Surgery*; grupo 2: *Journal of the American College of Surgeons, Surgery y Archives of Surgery*; y grupo 3: *The British Journal of Surgery y Annals of Surgery*. Los artículos fueron analizados de forma cruzada e independiente por 2 equipos de revisores.

Resultados. Un total de 751 artículos estaba relacionado con la terapéutica (41,2%). El FI promedio de los grupos en estudio fue $1,66 \pm 0,48$ para el grupo 1; $2,61 \pm 0,15$ para el grupo 2 y $4,69 \pm 1,57$ para el grupo 3 ($p < 0,001$). Se encontraron 16 revisiones sistemáticas, 82 ensayos clínicos aleatorios, 96 estudios de cohortes, 474 series de casos y 79 artículos de revisión. Sólo el 4,8% de los artículos era de grado de evidencia 1. Se verificaron diferencias significativas en el 15,1% de los estudios (revisiones sistemáticas

y ensayos clínicos, que fueron más frecuentes en el grupo 3; $p < 0,001$).

Conclusiones. Se verificó la asociación entre el FI de las revistas quirúrgicas y el grado de evidencia en publicaciones relacionadas con la terapéutica.

Palabras clave: *Bibliometría. Metodología de la investigación. Factor de impacto. Análisis de citación.*

IS IMPACT FACTOR AN APPROPRIATE INDEX TO DETERMINE THE LEVEL OF EVIDENCE OF STUDIES ON THERAPEUTIC PROCEDURES IN SURGERY JOURNALS?

Introduction. The aim of this study was to determine the association between the IF of surgery journals and the level of evidence of articles on therapeutic procedures published in these journals.

Material and method. A bibliometric study was performed. Journals listed in the 2002 ISI® under the subject heading of Surgery were included. Journals were classified by applying the 33rd percentile of their IF (group 1 from 0.128 to 1.690, group 2 from 1.691 to 2.886, and group 3 from 2.887 to 6.674). Simple randomized sampling was conducted of journals in each group (group 1: Eur J Surg, Am Surg and World J Surg; group 2: J Am Coll Surg, Surgery and Arch Surg; and group 3: Br J Surg and Ann Surg). A crossover analysis was performed by 2 independent teams of reviewers.

Results. There were 751 article on therapy (41.2%). The mean IF was 1.66 ± 0.48 for group 1, 2.61 ± 0.15 for group 2, and 4.69 ± 1.57 for group 3 ($P < .001$). There were 16 systematic reviews, 82 randomized clinical trials, 96 cohort studies, 474 case series and 79 review articles. Level 1 evidence was found in only 4.8% of the articles. Statistical differences were verified in 15.1% of the study sample, which was repre-

Estudio parcialmente financiado por proyecto DID-UFRO EP2102.

Correspondencia: Dr. C. Manterola.

Departamento de Cirugía. Universidad de La Frontera.

Casilla, 54-D. Temuco. Chile.

Correo electrónico: cmantero@ufro.cl

Manuscrito recibido el 8-3-2005 y aceptado el 15-3-2005.

sented by systematic reviews and randomized clinical trials (most frequent in group 3 ($P<.001$).

Conclusions. An association between the IF of scientific surgical journals and the level of evidence in articles published on therapy was verified.

Key words: *Bibliometrics. Research design/methods. Impact factor. Citation analysis.*

Introducción

Los análisis bibliométricos pueden ser útiles para instituciones y organizaciones. Sin embargo, su validez suele ser motivo de polémica, en especial el factor de impacto de las revistas (FI), tema de interés para autores y editores¹.

Es usual que las revistas más consultadas sean las de mayor FI, pues se supone que éstas son las que contienen los artículos de mejor calidad metodológica. El FI se calcula dividiendo el número de citaciones en el año 3 de cualquier ítem publicado por la revista en cuestión durante los años 1 y 2, por el número de artículos publicados en esa revista en ambos años².

Por otra parte, en las últimas 2 décadas se ha utilizado el concepto de "grados de evidencia" en la bibliografía científica, en especial en el ámbito del diagnóstico, la terapéutica, el pronóstico y los estudios económicos. De hecho, éste ha generado tal impacto que el concepto original descrito por Sackett³ ha ido evolucionando con los años, agregándose innovaciones y modificándose progresivamente hasta transformarse en una nueva generación de grados de evidencia⁴⁻⁸.

Nuestra hipótesis es que las revistas quirúrgicas con mayor FI concentran artículos relacionados con la terapéutica, en adultos, de mayor grado de evidencia, por lo que el FI constituiría un índice apropiado para determinar el grado de evidencia para el tratamiento y, por ende, sería un índice apropiado para comparar revistas en este ámbito.

El objetivo de este estudio es determinar la asociación entre el FI de revistas científicas quirúrgicas y el grado de evidencia de los artículos relacionados con la terapéutica publicados en ellas.

Material y método

Diseño

Estudio bibliométrico.

Selección de revistas

Mediante un muestreo aleatorio simple entre las publicaciones de los años 2000 al 2003 se eligió el año 2002 para realizar este estudio. Se incluyeron entonces todos los títulos relacionados con cirugía general, digestiva, cardiovascular, plástica y torácica, publicados en inglés listados en el ISI® Journal Citation Report del año 2002 (tabla 1). La mediana de FI para las revistas quirúrgicas ISI para ese año fue 1,48 (0,128-6,674). De acuerdo con estos valores y su distribución, las revistas fueron agrupadas aplicando el percentil 33 del FI de todas ellas, confor-

TABLA 1. Revistas quirúrgicas consideradas por título, ISSN y factor de impacto (n = 29)

Revista	ISSN	IF
<i>Annals of Surgery</i>	0003-4932	6,674
<i>Journal of Endovascular Surgery</i>	1074-6218	5,467
<i>British Journal of Surgery</i>	0007-1323	3,464
<i>Journal of Vascular Surgery</i>	0741-5214	3,145
<i>Journal Of Thoracic And Cardiovascular Surgery</i>	0022-5223	2,818
<i>Archives Of Surgery</i>	0004-0010	2,792
<i>Surgery</i>	0039-6060	2,615
<i>Journal Of The American College Of Surgeons</i>	1072-7515	2,429
<i>Surgical Endoscopy-Ultrasound And Interventional Techniques</i>	0930-2794	2,374
<i>Annals Of Thoracic Surgery</i>	0003-4975	2,141
<i>American Journal Of Surgery</i>	0002-9610	2,131
<i>Obesity Surgery</i>	0960-8923	2,099
<i>European Journal of Cardio-Thoracic Surgery</i>	1010-7940	1,676
<i>World Journal of Surgery</i>	0364-2313	1,644
<i>European Journal of Vascular And Endovascular Surgery</i>	1078-5884	1,484
<i>Plastic and Reconstructive Surgery</i>	0032-1052	1,436
<i>American Surgeon</i>	0003-1348	1,182
<i>Annals of Vascular Surgery</i>	0890-5096	1,06
<i>Thoracic and Cardiovascular Surgeon</i>	0171-6425	0,995
<i>British Journal of Plastic Surgery</i>	0007-1226	0,816
<i>European Journal of Surgery</i>	1102-4151	0,794
<i>Cardiovascular Surgery</i>	0967-2109	0,776
<i>Digestive Surgery</i>	0253-4886	0,744
<i>Australian and New Zealand Journal of Surgery</i>	0004-8682	0,723
<i>Annals of Plastic Surgery</i>	0148-7043	0,714
<i>Journal of Cardiovascular Surgery</i>	0021-9509	0,55
<i>Canadian Journal of Surgery</i>	0008-428x	0,503
<i>International Surgery</i>	0020-8868	0,386
<i>South African Journal of Surgery</i>	0038-2361	0,128

mándose de esta manera 3 grupos (grupo 1 de 0,128 a 1,690, grupo 2 de 1,691 a 2,886, y grupo 3 de 2,887 a 6,674). Se realizó entonces un muestreo aleatorio simple mediante asignación computacional para seleccionar las revistas para cada grupo, después de la cual los grupos quedaron conformados de la siguiente forma: grupo 1: *European Journal of Surgery*, *American Surgeon* y *World Journal of Surgery*; grupo 2: *Journal of the American College of Surgeons*, *Surgery* y *Archives of Surgery*, y grupo 3: *British Journal of Surgery* y *Annals of Surgery*. Se excluyeron revistas de revisión, como *Surgical Clinics of North American* y *Current Opinion in Surgery*, y revistas de investigación, como *Journal of Surgical Research* o *Journal of Investigative Surgery*. Se consideraron para el análisis todo tipo de diseños, excluyéndose editoriales, cartas al editor, investigación de laboratorio y estudios realizados en la población pediátrica. Entre los 5.748 títulos que componen el listado ISI® del Journal Citation Report del año 2002 se encontraron 53 títulos quirúrgicos que cumplían los criterios de inclusión. El FI más alto lo registró *Annals of Surgery*, que se encuentra en el puesto 142 del ranking general del listado ISI® con un FI de 6,674, y la revista quirúrgica con menor FI es *South African Journal of Surgery* con 0,128. De esta forma, se verificó que la mediana de FI de las revistas consideradas fue 2,429 (0,794 a 6,674).

Tamaño de la muestra

Se estimó una muestra de 8 revistas con base en el número de revistas quirúrgicas ISI, un intervalo de confianza del 99%, una frecuencia de esperada de asociación de 80% y un peor resultado de asociación de 20%. La muestra se distribuyó en los grupos antes definidos según el número de revistas disponibles en cada uno de esos intervalos en el momento del estudio.

TABLA 2. Grupos estudiados y diseños clínicos encontrados (n = 751) (p < 0,001)

Diseños	Grupo 1 (%)	Grupo 2 (%)	Grupo 3 (%)	Total (%)
RSL	0,3	1,5	6,0	2,1
ECA	7,6	9,1	18,9	10,9
Cohortes	14,2	10,6	13,5	12,8
Casos y controles	0,0	1,5	0,0	0,5
Serie de casos	71,9	61,7	50,8	63,2
Artículos de revisión	6,0	15,5	10,8	10,5
Total	100,0	100,0	100,0	100,0

RSL: revisiones sistemáticas de la bibliografía de ECA; ECA: ensayos clínicos aleatorizados.

TABLA 4. Grupos en estudio y grados de evidencia verificados (n = 751) (p < 0,001)

Grados de evidencia	Grupo 1 (%)	Grupo 2 (%)	Grupo 3 (%)	Total (%)
1	3,0	3,4	9,7	4,8
2	6,6	8,3	18,9	10,3
3	0,0	1,5	0,0	0,5
4	84,4	71,2	60,5	73,9
5	6,0	15,5	10,8	10,5
Total	100,0	100,0	100,0	100,0

RSL: revisiones sistemáticas de la bibliografía de ECA; ECA: ensayos clínicos aleatorizados.

TABLA 3. Grados de evidencia de los artículos analizados (n = 751)

Grado de evidencia	Número de artículos	%
1	36	4,8
2	77	10,3
3	4	0,5
4	555	73,9
5	79	10,5
Total	751	100,0

Metodología de análisis

Todos los artículos de terapia en adultos humanos publicados en las revistas seleccionadas fueron revisados y analizados aplicando la clasificación de Sackett modificada de grados de evidencia^{7,8}. La revisión fue realizada de forma cruzada e independiente por 2 equipos de revisores, de 2 revisores cada uno. Las discrepancias entre revisores y equipos de revisores fueron resueltas por consenso.

Definiciones

Para los propósitos de este estudio, el grado de evidencia 1 incluye revisiones sistemáticas (RSL) de ensayos clínicos aleatorios controlados (ECA) y ECA individuales de intervalo de confianza estrecho. El nivel 2 incluye RSL de estudios de cohortes, estudios de cohortes individuales y ECA individuales de baja calidad metodológica. El nivel 3 incluye RSL de estudios de casos y controles, y estudios de casos y controles individuales. El nivel 4 incluye serie de casos, estudios de cohortes y estudios de casos y controles de baja calidad. El nivel 5 incluye la opinión de expertos sin evaluación crítica explícita, representada por los artículos de revisión^{7,8}.

Plan de análisis

Después de un análisis exploratorio de los datos se aplicó la estadística descriptiva (cálculo de promedios, desviación estándar y medianas)

y analítica para la comparación de grupos (prueba de la χ^2 de Pearson, exacto de Fisher, ANOVA y no paramétricas). Se consideraron estadísticamente significativos los valores de $p < 1\%$.

Resultados

Durante el año 2002 se publicaron 1.825 artículos en las revistas seleccionadas. De éstas, 751 eran de terapia (41,2%); 302 artículos en el grupo 1 (40,2%), 264 artículos en el grupo 2 (35,2%) y 185 artículos en el grupo 3 (24,6%).

El FI fue $1,66 \pm 0,48$ para el grupo 1, $2,61 \pm 0,15$ para el grupo 2, y $4,69 \pm 1,57$ para el grupo 3 ($p < 0,001$).

Al aplicar la clasificación de Sackett se encontraron 16 RSL (2,1%), 82 ECA (10,9%), 96 estudios de cohortes (12,8%), 4 estudios de casos y controles (0,5%), 474 series de casos (63,2%) y 79 artículos de revisión (10,5%) (tabla 2). Por lo tanto, sólo el 4,8% de los artículos corresponde a la grado 1 de evidencia, y la mayor parte fueron clasificados en los grados 4 y 5 (tabla 3).

Se verificaron diferencias estadísticamente significativas en el 15,1% de la muestra, representada por RSL y ECA de calidad diversa (grados de evidencia 1 y 2). Estos diseños fueron más frecuentes en el grupo 3. Consecutivamente, las series de casos fueron menos prevalentes en el grupo 3 ($p < 0,001$) (tabla 4).

Discusión

Las revistas publican diferentes tipos de materias, muchas de las cuales no aportan investigación sustancial y, en este aspecto, la cirugía no es diferente. Por ello, se generó un algoritmo especial que se utiliza para calcular el número de ítems sustanciales de una revista. Éste es el FI, herramienta proporcionada por el Institute for Scientific Information (Philadelphia) y que se ha transformado en la herramienta más importante en el proceso de evaluación de la investigación científica y del trabajo académico; además, puede ser usada como un criterio de evaluación de la actividad de cada revista debido a que, en el "mercado de la investigación", el FI provee una evidencia cuantitativa para editores y editoriales para posicionar sus revistas en relación con la "competencia". Es un método de medición de calidad de las revistas, asumiendo que la citación de artículos entre las revistas se encuentra normalmente distribuida¹. Es más, el FI suele utilizarse para medir la calidad de artículos individuales, hecho erróneo, pues es un índice cuyo cálculo no se proyecta de forma transversal a los artículos de una revista.

Sin embargo, para otros, el FI es un indicador de citaciones y no de calidad; de hecho, se sugiere que no debe ser utilizado para valorar el trabajo de los autores¹⁰⁻¹³. De todos modos, estudios empíricos en relación con la validez del FI como indicador carecen de calidad, y se sabe que hay múltiples variables que pueden intervenir en el cálculo del FI y que pueden verse afectadas o incluso constituir sesgos como, por ejemplo, el tipo de revista y su peso, la lengua de la publica-

ción, el nombre de los autores, las autocitaciones, las citaciones repetidas (pocos artículos tienen muchas citaciones y la mayoría son citados de forma ocasional o nunca), la especialidad médica de la revista, etc. Por otra parte, la elección de las citaciones suele ser subjetiva y no es infrecuente observar la incoherencia de algunas citaciones. De hecho, un grupo de editores ha desarrollado una nueva propuesta para clasificar los artículos originales relacionados con ciencias biomédicas, medicina clínica y salud pública con la idea de evitar que al menos la lengua y las especialidades médicas sean un factor que pueda alterar la citación de un artículo¹³.

Nuestros hallazgos sugieren que el FI puede considerarse un indicador de calidad metodológica razonable para la evaluación de artículos de terapia en adultos humanos, publicados en las revistas estudiadas. Se constató una asociación entre el FI y grado de evidencia. Esta asociación se relaciona con los artículos de grado 1 y 2 (más prevalentes en el grupo 3). No obstante, más del 80% de los artículos de terapia publicados en las revistas en estudio en el período analizado tiene un grado de evidencia 4 y 5, y son más frecuentes en las revistas del grupo 1.

Bibliografía

1. Garfield E. Fortnightly review: how can impact factors be improved? *BMJ*. 1996;313:411-3.
2. Garfield E. Journal impact factor: a brief review. *Can Med Assoc J*. 1999;161:979-80.
3. Sackett DL. Rules of evidence and clinical recommendations on use of antithrombotic agents. *Chest*. 1986;89 Suppl 2:2S-3S.
4. Meakins JL. Innovation in surgery: the rules of evidence. *Am J Surg*. 2002;183:399-405.
5. Oxford Centre for Evidence-Based Medicine (accedido oct. 2003). Disponible en: http://www.cebm.net/levels_of_evidence.asp
6. Corporación EBM Colombia (accedido oct. 2003). Disponible en: http://www.ebmcolumbia.org/Curso_CBasicos/Niveles_de_Evidencia.htm
7. Manterola C. Medicina basada en la evidencia. Conceptos generales y razones para su aplicación en cirugía. *Rev Chil Cir*. 2002;54:550-4.
8. Manterola C, Pineda V, Vial M, Losada H, Muñoz S. Revisión sistemática de la literatura. Propuesta metodológica para su realización en cirugía. *Rev Chil Cir*. 2003;55:204-8.
9. Dean AG. Epi Info version 6: a work processing, database and statistics system for epidemiology on microcomputers. Atlanta: Centers for Disease Control. 1994; p. 399-401.
10. Saha S, Saint S, Christakis DA. Impact factor: a valid measure of journal quality? *J Med Libr Assoc*. 2003;91:42-6.
11. Ojasoo T, Maisonneuve H, Matillon Y. The impact factor of medical journals, a bibliometric indicator to be handled with care. *Presse Med*. 2002;31:775-81.
12. Golder W. The impact factor: a critical analysis. *Rofo*. 1998;169:220-6.
13. Reyes H. The "impact factor" and the impact of medical journals. *Rev Med Chil*. 1998;126:135-8.