

Confiabilidad de la escala MINCIR para valorar calidad metodológica de estudios de terapia*

Drs. JAVIER MORAGA C.¹, MARÍA E. BURGOS D.^{1,2,3}, CARLOS MANTEROLA D.^{1,2,3}, ANTONIO SANHUEZA C.^{1,2,4}, RICARDO CARTES-VELÁSQUEZ V.², SEBASTIÁN URRUTIA V.³
GRUPO MINCIR (METODOLOGÍA E INVESTIGACIÓN EN CIRUGÍA)

¹ Programa de Magíster en Ciencias Médicas, Mención Cirugía. Facultad de Medicina.

² Programa de Doctorado en Ciencias Médicas, Facultad de Medicina.

³ Departamento de Cirugía y Traumatología. Facultad de Medicina.

⁴ Departamento de Matemáticas y Estadística.

Universidad de La Frontera.

Temuco, Chile.

Abstract

Validation of a scale for assessing methodological quality of therapy studies. Pilot study

Introduction: The methodological quality (MQ) is a complex multidimensional concept that assesses multiple items as design, methodology and analysis. There are few tools to assess MQ of studies in the field of therapy and these are also partially validated. The aim of this study is to determine the intra and inter observer reliability of the MINCIR scale for assess MQ of therapy articles. **Methods:** Study of validation scales. Was calculate sample size considering confidence level of 95%, accuracy of 2 points and standard deviation of 4.95, which gives 21 articles, select by simple random probabilistic sampling with Stata 10.0. Was determined intra and inter observer reliability of the scale and each one of its domains using intraclass correlation coefficient (ICC). **Results:** The ICC for interobserver reliability was 0.9. The ICC observed for the domains one, two and three was 0.97, 0.88 and 0.64 respectively. The ICC observed for intraobserver reliability was 0.99. The ICC observed for the domains one, two and three were 1; 1 and 0.91 respectively. **Conclusion:** Interobserver and intraobserver reliability for the MINCIR scale to assess MQ of therapy articles were determined.

Key words: Scales, "Reproducibility of results" [MeSH], Methods [MeSH], Therapy studies, "Bibliometrics" [MeSH].

Resumen

Introducción: La calidad metodológica (CM) es un constructo multidimensional complejo que evalúa múltiples ítems como diseño, metodología y análisis realizados. Existen escasos instrumentos para valorar CM de estudios en el ámbito de la terapia y además estos se encuentran parcialmente validados. El objetivo de este estudio es determinar la confiabilidad intra e interobservador de la escala MINCIR para valorar CM en estudios de terapia. **Material y Método:** Diseño de validación de escalas. Se calculó un tamaño de muestra

*Recibido el 12 de noviembre de 2012 y aceptado para publicación el 17 de enero de 2013.

Los autores no refieren conflictos de interés.

Correspondencia: Dr. Carlos Manterola D.
Manuel Montt 112, Oficina 408. Temuco, Chile.
cmantero@ufro.cl

considerando un nivel de confianza de 95%, una precisión de 2 puntos, y una desviación estándar de 4,95; lo que arrojó 21 artículos. Estos fueron seleccionados mediante muestreo probabilístico aleatorio simple con programa Stata 10.0. Se determinó confiabilidad intra e interobservador de la escala y de cada uno de sus dominios mediante coeficiente de correlación intraclase (CCI). **Resultados:** La escala evaluada presenta un CCI para confiabilidad interobservador de 0,9; el CCI para el dominio uno, dos y tres es de 0,97; 0,88 y 0,64 respectivamente. El CCI para la confiabilidad intraobservador de la escala es de 0,99 y para los dominios uno, dos y tres es de 1; 1; y 0,91 respectivamente. **Conclusión:** Se determinó confiabilidad inter e intraobservador de la escala MINCIR para valorar CM de artículos de terapia.

Palabras clave: Escalas, confiabilidad, metodología, calidad metodológica, estudios de terapia, bibliometría.

Introducción

Durante mucho tiempo las decisiones en salud se tomaron basadas en la experiencia de las personas¹, la medicina basada en la evidencia que inicia su desarrollo en los años sesenta conducida por David Sackett, vino a cambiar esta conducta, intentando integrar las competencias clínicas individuales con la mejor evidencia clínica externa disponible². Es en este escenario que se hace indispensable contar con instrumentos de medición válidos y confiables^{3,4}.

La calidad metodológica (CM) de los artículos es un constructo multidimensional complejo que evalúa diferentes ítems como: diseño, tamaño de la muestra empleada, metodología, análisis realizados, calidad del reporte, etc⁵⁻⁷ (Figura 1).

Actualmente, existen algunos sistemas diseñados para la jerarquización de la evidencia disponible como los propuestos por Sackett^{8,9} y Oxford^{8,10}; estas corresponden a escalas alfanuméricas basadas en el diseño del estudio. Existen también recomendaciones para el reporte de resultados o “listas de chequeo” de diferentes tipos de diseño como son las iniciativas CONSORT¹¹, TREND¹², GRADE¹³, MOOSE¹⁴, QAREL¹⁵, QUADAS¹⁶, QUORUM¹⁷, REMARK¹⁸, STARD¹⁹ y STROBE²⁰. Tanto los niveles de evidencia como las recomendaciones para el reporte han sido utilizados para evaluar la CM de artículos publicados, sin embargo, estas no fueron diseñadas con ese objetivo y tampoco han sido validadas para tales fines.

A nivel local el grupo MINCIR posee una línea de investigación en CM, cuyo objetivo ha sido la generación de escalas para valorar CM en diferentes escenarios clínicos como son terapia, pronóstico y pruebas diagnósticas^{1,5,7}. Estas, han sido utilizadas para valorar CM de artículos de diferentes revistas²¹⁻²⁴ y para la realización de revisiones sistemáticas de la literatura con metodología alternativa²⁵⁻²⁷.

La escala para valorar CM de estudios de terapia (Figura 2) se encuentra parcialmente validada, contando con validez de contenido y de fachada; proceso que se pretende completar con validación de criterio, constructo y estudios de confiabilidad.

El objetivo de este estudio es determinar la confiabilidad intra e interobservador de la escala MINCIR para valorar CM de estudios de terapia.

Material y Método

Diseño: Estudio de validación de escalas.

Población a estudio: Artículos de terapia en humanos publicados en las revistas *Annals of Surgery*, *British Journal of Surgery*, *Archives of Surgery*, *European Journal of Surgery*, *Journal of American College of Surgeons*, *Surgery*, *World Journal of Surgery*, *American Surgeon*, *Cirugía Española* y *Revista Chilena de Cirugía*.

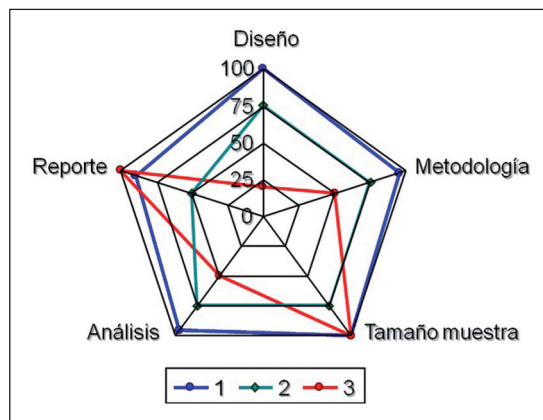


Figura 1. Concepto de multidimensionalidad de CM, representado en un gráfico polar, en el que se aprecia una pentágono que representa a diversos aspectos que se pueden valorar para definir el constructo CM. En color azul, se simula un artículo que cumple con casi un 100% de todos los dominios evaluados; en rojo, un artículo que cumple con un 75% en los dominios evaluados; y en verde, un artículo que cumple en un 100% con dos de los dominios, en un 50% con otros dos y sólo en un 25% en el dominio restante. Esta imagen permite entender mejor porque algunos artículos de buen número de casuística pueden tener inadecuada CM (mala calidad del reporte de los resultados, inadecuada metodología, etc.).

<i>Ítem 1. Diseño del estudio</i>	<i>Puntaje asignado</i>
Ensayo clínico multicéntrico	12
Ensayo clínico controlado, con asignación aleatoria y doble enmascaramiento*	9
Ensayo clínico sin asignación aleatoria, con enmascaramiento simple o sin enmascaramiento**	6
Cohorte concurrente o prospectiva	4
Cohorte histórica o retrospectiva y estudios de casos y controles	3
Corte transversal	3
Serie de casos y reportes de casos	1
<i>Ítem 2. Población estudiada por factor de justificación</i>	<i>Puntaje asignado</i>
≥ 201	6 ó 12
151-200	5 ó 10
101-150	4 ó 8
61-100	3 ó 6
≤ 30	1 ó 2
<i>Ítem 3. Metodología empleada</i>	<i>Puntaje asignado</i>
Objetivo	
- Se plantean objetivos claros y concretos	3
- Se plantean objetivos vagos	2
- No se plantean objetivos	1
Diseño	
- Se menciona y justifica el diseño empleado	3
- Se menciona el diseño empleado	2
- No se menciona ni justifica el diseño empleado	1
Criterios de selección de la muestra	
- Se describen criterios de inclusión y de exclusión	3
- Se describen criterios de inclusión o de exclusión	2
- No se describen criterios de selección	1
Tamaño de la muestra	
- Justifica la muestra empleada	3
- No justifica la muestra empleada	1
Score final	Puntaje final
Ítem 1 + (ítem 2 x factor de justificación) + ítem 3	6 a 36
*Incluye ensayos clínicos con asignación aleatoria restringida (conglomerados) y cuasi-experimentales.	
** Incluye estudios experimentales (antes y después)	

Figura 2. Escala MINCIR de CM de estudios de terapia.

Muestreo: Para la selección del año, revistas y artículos se utilizó un muestreo probabilístico aleatorio simple lo que generó una base de datos compuesta por 226 artículos. A partir de esta base de datos se realizó un muestreo probabilístico aleatorio simple de 28 artículos con los que se llevó a cabo el estudio.

Criterios de selección

- **Criterios de inclusión:** Artículos de terapia en humanos, en sujetos mayores de 18 años, sin restricción de género ni área temática; relacionados con cirugía y disciplinas afines; publicados en revistas quirúrgicas indizadas en la base de datos Science Citation Index Expanded Web of Knowledge de Thomson Reuters.
- **Criterios de exclusión:** Editoriales, cartas al editor, artículos de revisión y discusión, guías clínicas, revisiones sistemáticas de la literatura, artículos de investigación *in vitro* y estudios biométricos.

Estimación del tamaño de la muestra: Se realizó el cálculo considerando un nivel de confianza del 95%, una precisión de 2 puntos y una desviación estándar de 4,95; lo que arrojó un tamaño de muestra mínimo de 21 artículos.

Maniobra: Se realizó estandarización entre los evaluadores y posterior aplicación de la escala MINCIR de terapia a los artículos seleccionados por parte de dos investigadores independientes (JM y MEB).

Variables en estudio

- **Confiabilidad interobservador:** Corresponde al grado de acuerdo que existe entre dos investigadores independientes al aplicar la escala a los mismos artículos en un mismo período de tiempo.
- **Confiabilidad intraobservador:** Corresponde al grado de acuerdo obtenido por el mismo investigador al aplicar la escala a los mismos artículos en dos momentos diferentes.

Plan de análisis: Después de realizar un análisis exploratorio de los datos se aplicó estadística descriptiva con cálculos de porcentajes. Posteriormente se realizó determinación de la confiabilidad interobservador e intraobservador mediante coeficiente de correlación intraclase (CCI), de la escala en su totalidad y de cada uno de sus dominios en particular.

Aspectos éticos: Se resguardó la confidencialidad de los autores de los artículos y de las revistas en las cuales los artículos fueron publicados. Por otra parte, los autores declaran no tener conflictos de interés con los resultados obtenidos.

Resultados

La base de datos generada inicialmente, quedó compuesta por 226 artículos, 28 de los cuales se utilizaron para realizar este estudio. De estos, el 28,6% son series de casos retrospectivas; 25,0% reportes de casos; 17,9% series de casos prospectivas; 14,3% cohortes retrospectivas; 7,1% cohortes prospectivas y 7,1% ensayos clínicos.

Tabla 1. Características de las publicaciones analizadas

Características	%	Valor
Área temática		
Partes blandas, plástica, hernia y mama	25,0	7
Esófago, estómago, intestino delgado, bazo y obesidad	17,8	5
Cardiovascular	14,3	4
Hígado, vía biliar y páncreas	14,3	4
Otras	28,6	8
Procedencia de los artículos		
América del Norte	32,2	9
Europa	32,2	9
América del Sur	17,8	5
Asia	10,7	3
Oceanía	7,1	2
Revistas de publicación de artículos		
British Journal of Surgery	17,8	5
Revista Chilena de Cirugía	17,8	5
Annals of Surgery	10,7	3
Surgery	10,7	3
Archives of Surgery	10,7	3
European Journal of Surgery	10,7	3
American Surgeon	10,7	3
Cirugía Española	7,2	2
Journal of American College of Surgeons	3,7	1

El 85,7% de los estudios son de nivel de evidencia 4 y el restante 14,2% de nivel de evidencia 2b; representan de forma proporcionada a diversas áreas temáticas de la disciplina, así como también a lugares de origen. Las características generales de los artículos analizados se pueden observar en la Tabla 1.

El promedio y la mediana de puntuación de CM de los artículos analizados fueron de $10,3 \pm 4,1$ y de 9 puntos respectivamente (intervalo de confianza del 95%: 8,6 a 11,9); y con valores extremos de 6 y 23 puntos. La distribución de la CM de los artículos analizados se puede observar en la Figura 3.

La confiabilidad interobservador de la escala tiene un CCI de 0,9; y para cada uno de sus dominios (1, 2, 3), es de 0,97; 0,88; y 0,64 respectivamente.

La confiabilidad intraobservador de la escala tiene un CCI de 0,99; y para cada uno de sus dominios (1, 2, 3) es de 1; 1 y 0,91 respectivamente.

Discusión

Hoy en día se hace cada vez más necesario contar con instrumentos válidos y confiables para afrontar la práctica clínica así como también para enfrentar la conducción de estudios. Es por lo anterior, que nos surge la necesidad de seguir validando la escala MINCIR para evaluación de CM de estudios de terapia. Dentro de este proceso de validación, este estudio piloto constituye el primer paso, al analizar la confiabilidad tanto interobservador como intraobservador de la escala.

El desempeño de la escala en diferentes escenarios de CM se probó aplicándola en artículos

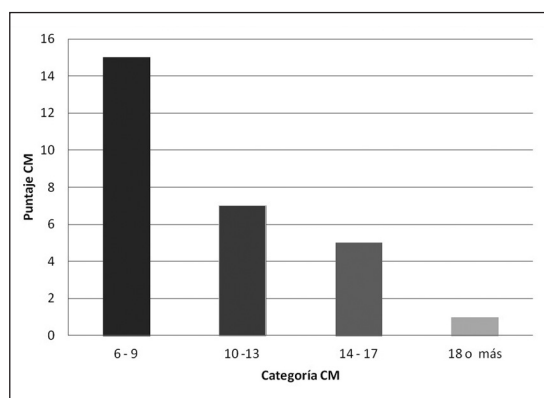


Figura 3. Distribución de la CM de los artículos analizados, agrupados en categorías por el puntaje obtenido al aplicar la escala MINCIR. Es importante señalar que la mayor parte de los estudios se agrupan en los 6 y 13 puntos (es decir, de CM pobre); y sólo algunos pocos se acercan a los 18 puntos o lo superan (punto de corte para el constructo).

provenientes de revistas de distinto valor de factor de impacto (desde 0,18 a 7,49).

Al realizar el estudio pudimos darnos cuenta que para la aplicación de cada instrumento se hace indispensable realizar un proceso de calibración entre los usuarios de la herramienta; ya que, de no realizarse, se corre el riesgo de tener criterios de aplicación diferentes y por ende resultados dispares, lo que constituiría un error sistemático de su aplicación y por ende un sesgo de medición.

Al realizar el análisis exploratorio de los datos, es interesante destacar que los porcentajes de cada uno de los diseños observados en la muestra se condicen fielmente con la distribución de los distintos tipos de diseños de investigación observados al realizar estudios bibliométricos²¹⁻²⁴.

En esta segunda etapa de la validación de la escala MINCIR terapia, pudimos darnos cuenta que ésta presenta un buen comportamiento a nivel de confiabilidad inter e intraobservador pues en todas las mediciones, salvo una; se verificaron valores superiores a 0,71; y en varias superiores a 0,90; lo que se interpreta como valores con fuerza de concordancia “buenos” y “muy buenos”²⁸.

En resumen, se puede concluir señalando que los resultados obtenidos nos dan luces para continuar con el proceso final de validación; contribuyendo de este modo a generar conocimiento en el ámbito de la medicina basada en evidencia y la investigación en metodología; aportando instrumentos de evaluación de CM de aplicación fácil y universal.

Referencias

1. Losada H, Manterola C, Pineda V, Vial M, Sanhueza A, Grupo MINCIR. Diseño de una escala para la evaluación de calidad metodológica de estudios de pronóstico. *Rev Chil Cir.* 2009;61:59-72.
2. Sackett D, Roseberg WM, Muir Gray JA, Haynes RB, Richardson WS. Evidence based medicine: What it is and what it isn't. *Br Med J.* 1996;312:71-2.
3. Manterola C. El proceso de medición con variables cualitativas y su aplicación en cirugía. *Rev Chil Cir.* 2002;54:307-15.
4. Alarcón A, Muñoz S. Medición en salud: Algunas consideraciones metodológicas. *Rev Med Chile* 2008; 136:125-30.
5. Manterola C, Pineda V, Vial M, Losada H, Muñoz S. Revisión sistemática de la literatura. Propuesta metodológica para su realización. *Rev Chil Cir.* 2003;55:204-8.
6. Manterola C. Respecto de la calidad metodológica de los artículos que se publican en las revistas biomédicas. *Rev Chil Cir.* 2005;57:449-50.
7. Burgos M, Manterola C, Sanhueza A. Diseño de una escala para evaluar calidad metodológica de estudios de pruebas diagnósticas. Estudio piloto. *Rev Chil Cir.* 2011;63:493-97.
8. Manterola C, Zavando D. Cómo interpretar los “Niveles de Evidencia” en los diferentes escenarios clínicos. *Rev Chil Cir.* 2009;61:582-95.
9. Sackett DL, Wennberg JE. Choosing the best research design for each question. *BMJ* 1997;315(7123):1636.
10. Oxford Centre for Evidence-based Medicine (CEBM). Centre for Evidence Based Medicine - Levels of Evidence (March 2009). Available from: <http://www.cebm.net/index.aspx?o=1025>. Visitado el 22 de junio de 2009.
11. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637-9.
12. Don C, Des Jarlais, Cynthia Lyles, Nicole Crepaz, and the TREND Group. Improving the Reporting Quality of Nonrandomized Evaluations of Behavioral and Public Health Interventions: The TREND Statement. *Am J Public Health* 2004;94:361-6.
13. Guyatt G, Oxman A, Kunz R, Falck-Ytter, Vist G, Liberati A, et al. GRADE: Going from evidence to recommendations. *BMJ* 2008;336:1049-51.
14. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting. *JAMA* 2000;283:2008-12.
15. Lucas N, Macaskill P, Irwig L, Bugduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *Journal of Clinical Epidemiology* 2010;63:854-61.
16. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 2003;3:25.
17. Moher D, Cook D, Eastwood S, Olkin I, Rennie D, Stroup D. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUORUM statement. *British Journal of Surgery* 2000;87:1448-54.
18. McShane L, Altman D, Sauerbrei W, Taube S, Gion M, Clark G for the Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics. Reporting recommendations for tumour marker prognostic studies (REMARK). *British Journal of Cancer* 2005;93:387-91.
19. Bossuyt P, Reitsma J, Bruns D, Gatsonis C, Glasziou P, Irwig L, et al, for the STARD group. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Family Practice* 2004;21:4-10.
20. Vandenberg J, Von Elm E, Altman D, Gotzsche P, Mulrow C, Pocock S, et al, for the STROBE Initiative. Strengthening The Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLoS Medicine* 2007;4:e297.

21. Pineda V, Manterola C, Vial M, Losada H. ¿Cuál es la calidad metodológica de los artículos referentes a terapia publicados en la Revista Chilena de Cirugía? *Rev Chil Cir.* 2005;57:500-7.
22. Manterola C, Pineda V, Vial M, Losada H, and the MINCIR group. What Is the Methodologic Quality of Human Therapy Studies in ISI Surgical Publications? *Ann Surg.* 2006;244:827-32
23. Manterola C, Busquets J, Pascual M, Grande L. What is the methodological quality of articles on therapeutic procedures published in *Cirugía Española*? *Cir Esp.* 2006;79:95-100.
24. Manterola C, Grande L. Methodological quality of articles on therapeutic procedures published in *Cirugía Española*. Evaluation of the period 2005-2008. *Cir Esp.* 2010;87: 244-50.
25. Vial M, Manterola C, Pineda V, Losada H. Coledocolitiasis. Elección de una terapia basada en la evidencia. Revisión sistemática de la literatura. *Rev Chil Cir.* 2005;57:404-11.
26. Manterola C, Pineda V, Vial M. Resección abierta frente a laparoscópica en el cáncer de colon no complicado. Revisión sistemática. *Cir Esp.* 2005;78:28-33.
27. Manterola C, Pineda V, Vial M, Losada H, Muñoz S. Surgery for Morbid Obesity: Selection of Operation Based on Evidence from Literature Review. *Obs Surg.* 2005;15:106-13.
28. Pita Fernández S, Pértegas Díaz S. La fiabilidad de las mediciones clínicas: el análisis de concordancia para variables numéricas. http://www.fisterra.com/mbe/investiga/conc_numerica/conc_numerica.asp. Consultado el 30 de octubre de 2012.