

## Cómo valorar e interpretar un artículo sobre tratamiento o procedimientos terapéuticos\*

Dr. CARLOS MANTEROLA D.<sup>1,2</sup>, QF. MARCELA CORTÉS J.<sup>2,3</sup>

<sup>1</sup> Departamento de Cirugía y Traumatología, Universidad de La Frontera.

<sup>2</sup> Programa de Doctorado en Ciencias Médicas, Universidad de La Frontera.

<sup>3</sup> Universidad Católica de la Santísima Concepción.  
Temuco, Chile.

### Abstract

#### How to evaluate and interpret an article about therapy or therapeutic procedures

Independent of the thematic area and design employed, it was verified that between 40% and 60% of published studies in surgical scientific journals are related with treatment or therapeutic procedures (TP). The aim of this manuscript is to present a synthesis of the key elements for proper writing, reading and assessment of articles related with treatment or TP, regardless of disciplinary area which has arisen in the investigation. Who writes a manuscript relating to treatment or TP, or who read an article of this type should require at least a clear, accurate and concise objective, on the stage of the research being conducted, the explicit mention of the design used with the respective methodological details inherent in this, and the endorsement and implementation of statistical tools and measures of association, or at least the numbers needed to calculate these values.

**Key words:** “Therapeutics” [Mesh], “Therapy” [Subheading], Therapeutic procedures, Treatment, “Cohort Studies” [Mesh] “Clinical Trial” [Publication Type], “Review” [Publication Type], “Meta-Analysis” [Publication Type], “Risk” [Mesh].

### Resumen

Independiente del área temática y diseño empleado, se ha verificado que entre el 40% y 60% de los estudios publicados en revistas científicas del ámbito quirúrgico corresponden a artículos de tratamiento o procedimientos terapéuticos (PT). El objetivo de este manuscrito es presentar una síntesis de los elementos fundamentales para una correcta escritura, lectura y valoración de este tipo de artículos, independiente del área disciplinaria en la que tenga origen la investigación realizada. Quien escribe un manuscrito relacionado con tratamiento o PT, o quien lee un artículo de este tipo debe exigir al menos un objetivo claro, preciso y conciso respecto del escenario de la investigación que se realizó; la mención explícita del diseño empleado con los respectivos detalles metodológicos inherentes a este; y la mención y ejecución de herramientas estadísticas y medidas de asociación, o al menos los números necesarios para poder calcular estos valores.

**Palabras clave:** Terapéutica, terapia, procedimientos terapéuticos, tratamiento, estudio de cohortes, ensayo clínico, revisión sistemática de la literatura, meta-análisis, riesgo.

\*Recibido el 29 de Junio de 2010 y aceptado para publicación el 27 de Julio de 2010.

Parcialmente financiado por proyecto DI09-0060 de la Dirección de Investigación Universidad de La Frontera.

Correspondencia: Dr. Carlos Manterola D.  
Casilla 54-D, Temuco, Chile. Fax: 56-45-325761.  
E-mail: cmantero@ufro.cl

## Introducción

Los artículos de investigación clínica se pueden agrupar según el tipo de escenario del que se trate, o de la pregunta de investigación que se intente responder. Así, tenemos artículos de tratamiento, prevención y rehabilitación, daño y etiología; artículos de pronóstico; de diagnóstico; de prevalencia y diagnóstico diferencial; y artículos de análisis económicos<sup>1,2</sup>.

En un manuscrito anterior, se esbozaron los conceptos necesarios para evaluar y comprender un artículo de diagnóstico. En esta oportunidad, delinearemos las nociones esenciales para valorar e interpretar un artículo relacionado con tratamiento o procedimientos terapéuticos (PT). Esto, considerando la alta frecuencia de este tipo de publicaciones; pues independiente del área temática y diseño empleado para su realización, se ha verificado que entre el 40% y 60% de los estudios publicados en revistas científicas del ámbito quirúrgico corresponden a artículos de tratamiento o PT<sup>3-6</sup>; razón más que suficiente para fundamentar una correcta lectura y valoración de estos.

Idealmente, el reporte de resultados de estudios de tratamiento o PT deberían surgir de estudios válidos y confiables; de un buen nivel de evidencia y grado de recomendación; es decir, a partir de revisiones sistemáticas de la literatura (RS) y de ensayos clínicos individuales (EC) controlados, con asignación aleatoria y enmascaramiento e intervalos de confianza estrechos; diseños que representan niveles de evidencia 1a y 1b respectivamente y grado de recomendación A<sup>2,7</sup>. No obstante esto, la realidad es diferente; y las publicaciones sobre tratamiento y PT abarcan una amplia variedad de formas y fondo; formas por la diversidad de diseños existentes, que van desde los clásicos observacionales hasta los experimentales; y fondo porque a pesar de encontrar en la mayoría de las publicaciones una mayor o menor aproximación a un diseño, es frecuente también encontrar en ellos debilidades que amenazan la validez y confiabilidad de sus resultados. Es así como se ha determinado, que entre un 70% y 80% de los artículos referentes a tratamiento o PT corresponden a estudios de nivel de evidencia tipo 4 (reportes y series de casos; y estudios de cohortes retrospectivas); con una calidad metodológica baja, que obedece a graves deficiencias metodológicas<sup>3-6</sup>. Datos que refuerzan aún más la idea de valorar de forma apropiada los artículos científicos a través de la lectura crítica de ellos; para lo que se dispone de herramientas específicas.

Los puntos clave de un manuscrito relacionado con tratamiento o PT son el título, la pregunta de investigación, el objetivo del estudio, el diseño

empleado y el nivel de evidencia respectivo; las herramientas estadísticas y las medidas de asociación utilizadas. Finalmente, algunos sistemas de puntuación y listas de comprobación, que corresponden a herramientas que ayudan en la valoración de artículos cuyo objetivo es reportar resultados en el ámbito de la terapia o PT.

El objetivo de este artículo es entregar conceptos metodológicos básicos que se han de considerar cuando se valora e interpreta un estudio sobre tratamiento o PT.

## El título

Es imprescindible un título sugerente, que induzca a la curiosidad del lector para que se anime a leer el resumen y luego el texto. Su función principal es describir con exactitud el contenido del manuscrito. Con el menor número de palabras posible debe poder esbozar la incertidumbre existente respecto del tema en cuestión; al mismo tiempo debe explicar el tipo de estudio<sup>8</sup>.

En algunas ocasiones, los autores seleccionan un título que cumple con estas características. Esto ocurre especialmente cuando la escritura se guía por algunos estándares especiales como CONSORT, QUOROM, STROBE, TREND, etc<sup>9-13</sup>. Por ejemplo, "Coledocolitiasis: Elección de una terapia basada en la evidencia. Revisión sistemática de la literatura"<sup>14</sup>; o "Resección abierta vs Laparoscópica en cáncer de colon no complicado"<sup>15</sup>. En ambos casos, se enuncia de forma más o menos implícita la pregunta de investigación y el objetivo del estudio; así como el diseño empleado (una RS). "Ensayo aleatorizado sobre división versus no-división de vasos cortos gástricos durante la funduplicatura de Nissen laparoscópica: resultados a 10 años"<sup>16</sup>. En este ejemplo, se expresa la pregunta de investigación y el objetivo del estudio; así como el diseño empleado (un EC).

En otras ocasiones, los autores optan un título que cumple parcialmente con estas características. "Eficacia de la funduplicatura de Nissen versus tratamiento médico en la regresión de la displasia de bajo grado en pacientes con esófago de Barrett: un estudio prospectivo"<sup>17</sup>; o, "Supresión ácida intraesofágica en enfermedad por reflujo gastroesofágico complicado:esomeprazol versus lansoprazol"<sup>18</sup>. En ambos ejemplos, se expresa de forma más o menos precisa la pregunta y el objetivo del estudio; no así el diseño empleado, que podría corresponder a una RS, un EC, un estudio de cohorte prospectivo o incluso retrospectivo.

Pero, lo más común es que se opte por un título simple; en el que no queda claro lo que los autores pretenden reportar. Por ejemplo "Tratamiento de en-

fermedad hidatídica del hígado rota a la vía biliar<sup>19</sup>”; o “Ruptura intrabiliar oculta de quistes hidatídicos del hígado<sup>20</sup>”. En ambos casos el título es una mera descripción de un fenómeno; por ende no hay forma de suponer que tipo de diseño se empleó. En el primer ejemplo al menos se menciona la palabra tratamiento, sin especificar de qué se trata; pero en el segundo es imposible suponer incluso que se trata de un artículo sobre tratamiento o PT.

### Pregunta de investigación y objetivos

A la hora de decidir si el estudio es de tratamiento o PT, la lectura de la pregunta de investigación, cuando el autor la reporta, entrega la información adecuada. De no ser reportada (lo más común), el objetivo del estudio puede ayudar a comprender la naturaleza o escenario clínico del artículo.

Respecto de la pregunta clínica, se debe considerar que su planteamiento en forma estructurada, es el primer paso para diseñar una investigación. Aunque en el manuscrito que se envía a publicación puede no ser imprescindible, si se expone debe ser ante todo precisa y enfocada al problema planteado. Para ello, la mejor forma de exponerla es cautelando la incorporación de una serie de componentes básicos:

1. El Problema que genera incertidumbre en un paciente o una población.
2. La Intervención que se desea valorar.

3. El Comparador para la intervención en estudio; que según la pregunta y el diseño empleado, se debe considerar.
4. La Variable resultado (*Outcome*) que se medirá a partir de la intervención aplicada.

Por ejemplo, en el caso que se quiera evaluar la efectividad de la Gastrectomía y linfadenectomía regional D<sub>1</sub> asociada a quimioradioterapia para el tratamiento del cáncer gástrico resecable, deberemos describir con suficiente detalle la población a estudio, la intervención, el comparador o tratamiento alternativo, y el período de tiempo si fuese necesario (Tabla 1).

Un recurso útil para recordar los componentes cuando se estructuran las preguntas clínicas es el de formular el problema a partir del esquema denominado PICoR.

Este sistema, permite convertir la necesidad de conocimiento en una pregunta específica de cuatro componentes: Paciente, Intervención, Comparación y Resultado (*Outcome*). Por ello es que es muy común referirse al sistema PICoR de elaboración de preguntas clínicas. Aunque los componentes “intervención” y “comparación” sugieren el carácter experimental del ensayo clínico y parecen limitar el sistema a escenarios sobre tratamiento o PT; puede recurrirse a este también para elaborar preguntas sobre etiología, diagnóstico y pronóstico<sup>21</sup> (Tabla 1).

Tabla 1. Sistema PICOR para generación de preguntas de investigación con algunos ejemplos

P (paciente)	I (intervención)	Co (comparador)	R (resultado o <i>outcome</i> )
Pacientes con esofagitis por reflujo gastroesofágico a repetición	Funduplicatura de Nissen vía laparoscópica	Administración de Esomeprazol a permanencia	<ul style="list-style-type: none"> <li>✓ Control de síntomas</li> <li>✓ Control de la aparición de lesiones erosivas de la mucosa esofágica</li> </ul>
¿Cuál es el mejor tratamiento para el control de síntomas de enfermedad por reflujo gastroesofágico y el desarrollo de esofagitis secundaria entre Nissen laparoscópico y Esomeprazol?			
Pacientes con hidatidosis hepática única, no complicada, de diámetro inferior a 7 cm	Administración de Albendazol	Periquistectomía laparoscópica	<ul style="list-style-type: none"> <li>✓ Recidiva</li> <li>✓ Morbilidad</li> <li>✓ Costes involucrados</li> </ul>
¿Cuál es el mejor tratamiento para la hidatidosis hepática única, no complicada, de diámetro inferior a 7 cm en términos de recidiva, entre la administración de Albendazol y la periquistectomía laparoscópica?			
Pacientes con cáncer gástrico resecable	Gastrectomía y linfadenectomía regional D <sub>1</sub> asociada a quimioradioterapia	Gastrectomía y linfadenectomía regional D <sub>2</sub>	<ul style="list-style-type: none"> <li>✓ Supervivencia a 5 años</li> <li>✓ Morbilidad</li> <li>✓ Mortalidad</li> <li>✓ Efectos adversos</li> <li>✓ Recidiva</li> <li>✓ Costes involucrados</li> </ul>
¿Cuál es el mejor tratamiento para el cáncer gástrico resecable en términos de supervivencia a 5 años, entre Gastrectomía y linfadenectomía regional D <sub>1</sub> asociada a quimioradioterapia y Gastrectomía y linfadenectomía regional D <sub>2</sub> ?			

En la situación antes expuesta, la pregunta podría redactarse como: ¿Cuál es el mejor tratamiento para el cáncer gástrico reseccable en términos de supervivencia a 5 años, entre Gastrectomía y linfadenectomía regional D<sub>2</sub> y Gastrectomía y linfadenectomía regional D<sub>1</sub> asociada a quimiorradioterapia?

La falta de claridad y precisión de una pregunta se asocia entre otras cosas a una alta probabilidad de error en el cálculo del tamaño de muestra para la conducción del estudio; y con ello, a la seguridad de la estimación de la muestra, de la precisión de la inferencia, del poder estadístico o la capacidad para detectar diferencias si es que existen, etc.

Por otra parte, el objetivo es el eje en torno al cual se construye la estructura del estudio. Si este no es claro, preciso y conciso, será difícil discernir sobre el tipo de estudio que se tiene enfrente; es más, en este tipo de situaciones (lamentablemente muy frecuentes), la redacción del objetivo sólo agregará mayor incertidumbre y dudas respecto de la selección de la población a estudio, del tamaño de la muestra necesario, de las variables a estudiar y el ulterior análisis estadístico. Un problema frecuente en los artículos biomédicos es que los objetivos de la investigación suelen ser vagos e inexactos; o incluso en ocasiones no constan en el manuscrito<sup>3-5</sup>. De tal forma que objetivos como “evaluar los resultados de la resección hepática por carcinoma hepatocelular de lóbulo caudado a corto y largo plazo”<sup>22</sup>; poco precisos, presentan el inconveniente de no dejar en claro qué resultados se van a evaluar (¿supervivencia global o libre de enfermedad?, ¿recurrencia?, ¿morbilidad?, etc); tampoco si se trata de pacientes con carcinoma hepatocelular en general o de algún subtipo o estadio de avance de la enfermedad; o incluso del tipo de resección que se realizó (¿parcial, total o extendida?). Una opción de mejora de esta situación podría ser: “evaluar los resultados de la lobectomía total del caudado en pacientes con carcinoma hepatocelular estadio II y Child Pugh A, en términos de supervivencia global y recurrencia”. Esto, debido a que, en relación a este ejemplo, habitualmente se valoran pacientes con distintos tipos de histología, estadios, reserva funcional hepática, tipo de resecciones, etc, y a pesar de todo ello, se sacan conclusiones que pueden aplicar para unos escenarios pero no para todos.

### **Tipos de diseño y su nivel de evidencia**

Considerando el planteamiento primordial de la pregunta sobre tratamiento o PT ¿Cómo cambia el tratamiento el curso clínico de la enfermedad?, se puede suponer que el responderla involucra entre una serie de variables a considerar, además al tiempo;

es decir, al período de seguimiento desde que se administra el tratamiento en valoración hasta que se produce un cambio en el curso clínico de la enfermedad.

Cualquier artículo debe declarar de forma explícita el diseño utilizado en la conducción del estudio, y los artículos referentes a tratamiento o PT no son una excepción.

Los diseños de estudio que involucran un tiempo de seguimiento son los estudios de cohortes (entre los estudios de tipo observacional) y los EC (entre los estudios experimentales); ambos con todas sus respectivas variantes<sup>23-26</sup>. No obstante ello, se ha verificado que entre el 70% y 80% de los artículos sobre tratamiento o PT corresponden a reporte y series de casos, prospectivas y retrospectivas<sup>3-5,27</sup>.

Ahora bien, si volvemos a la pregunta ¿cómo cambia el tratamiento el curso clínico de la enfermedad?, implícitamente se constata la siguiente interrogante ¿es el tratamiento en evaluación el causante del cambio de curso clínico de la enfermedad? En otras palabras ¿existe una asociación causal entre el tratamiento y la enfermedad? Se habla entonces de la relación Causa y Efecto. En este aspecto, se puede definir una asociación como la dependencia estadística existente entre dos o más factores, donde la ocurrencia de un factor aumenta (o disminuye) a medida que varía el otro. Pero su presencia, no significa que la relación observada sea necesariamente de causa-efecto, entonces el objetivo principal al valorar un estudio sobre tratamiento o PT es juzgar cuándo una relación tratamiento-enfermedad es causal.

Una asociación causal es aquella donde el cambio en la frecuencia y calidad de un tratamiento o PT genera como resultado un cambio correspondiente en la frecuencia de la enfermedad. De este modo, el juzgar cuando la asociación es causal se extiende más allá de la validez de los resultados de cualquier estudio e incluye la consideración de la información epidemiológica así como la credibilidad biológica de la hipótesis.

Si en un estudio sobre tratamiento o PT, se determina que el azar, el sesgo y la confusión son improbables para la explicación del cambio en el curso de la enfermedad, entonces se puede concluir que hay una asociación estadística válida. Es entonces necesario considerar si la relación es de causa y efecto, ya que la presencia de una relación estadísticamente válida no implica una relación de causalidad.

Hay criterios que pueden ayudar en los juicios de causalidad, incluyendo la fuerza de asociación, la credibilidad biológica de la hipótesis, la consistencia de los hallazgos así como otra información relacionada con la secuencia temporal y la presencia de una relación dosis-efecto<sup>28-30</sup>. El razonamiento básico para establecer una relación causal es la secuencia

de eventos, es decir, que la causa está presente antes que se produzca el efecto. Pero antes de establecer que dos o más factores tienen relación causa-efecto es necesario demostrar que la asociación entre ellos es válida; esto significa que una asociación válida es una asociación real o verdadera, donde el efecto del azar y del sesgo, es mínimo.

Sin embargo, se ha de considerar el error aleatorio o azar, pues este es inherente a todas las observaciones y se puede valorar aplicando una prueba de significación estadística cuyo objetivo es el valor de  $p$ <sup>31</sup>.

Otro instrumento para valorar la influencia del azar es la determinación del intervalo de confianza a 95% (IC 95%); que en términos estadísticos corresponde al intervalo de valores numéricos en el que se encuentra el valor poblacional que se está estimando con un nivel de confianza de 95% (sinónimo: estimación por intervalo)<sup>32</sup>. Así, el IC 95% de una medida de asociación contiene, con 95% de seguridad, el valor real de esta medida. Sin embargo, es relevante recordar que la información entregada por el valor de  $p$  y el IC 95% es complementaria. También es importante recordar que tanto el valor de  $p$  como el IC 95% son dependientes del tamaño de la muestra, de manera que a menor tamaño de muestra, mayor es el valor de  $p$  y más amplio el intervalo<sup>31,33</sup>.

Otro concepto que es necesario mencionar es el referente a los sesgos errores sistemáticos e involuntarios. Estos, se presentan cuando los componentes comparados no son lo suficientemente similares. Por ende, pueden ocurrir en cualquier fase del proceso de evaluación de una asociación; destacando los sesgos de selección, medición y confusión<sup>34,35</sup>.

En resumen, los efectos del azar y el sesgo en la evaluación de una asociación están relacionados con la calidad metodológica del estudio. Los tipos de sesgo considerados provocan una comparación distorsionada dentro del modelo de razonamiento causa-efecto. A pesar de una buena comparación interna, los resultados pueden no ser generalizables o extrapolados a un escenario distinto; esto ocurre cuando los grupos en estudio tienen una selección distorsionada de la población que supuestamente representan.

Por lo anteriormente expuesto, se puede señalar que el nivel de evidencia de los diseños de investigación clínica se encuentra directamente relacionado con la fuerza y tamaño de la asociación causal, en el entendido que estos nos informan de la proximidad con el valor real de la estimación. Desde este punto de vista, el mejor nivel de evidencia para estudios sobre tratamiento o PT lo dan las RS con o sin meta-análisis de EC<sup>36</sup>, seguido por los EC individuales de intervalo estrecho y los estudios observacionales (estudios de cohortes, de casos y controles, series

de casos)<sup>37</sup>.

Es así que los EC es fundamental que hayan sido planificados con asignación aleatoria y enmascaramiento. La ventaja del proceso de asignación aleatoria es que las variables relacionadas al pronóstico, conocidas y desconocidas, se distribuyen de manera similar entre los grupos en estudio, de tal modo que cualquier diferencia que se registre podrá atribuirse a las diferentes modalidades de tratamiento recibidas por uno y otro grupo. La ventaja del proceso de enmascaramiento es que permite una medición más objetiva, de suerte que ni el investigador ni el sujeto en estudio sepan cuál es la intervención que está asignada a cada grupo evitando así un sesgo por cualquiera de estos dos; situación que en muchos casos no es posible de realizar, ya sea por problemas éticos como de factibilidad. Un ejemplo típico de esta situación es cuando se intenta comparar resultados de cirugía laparoscópica y cirugía convencional. En estos casos el investigador no puede evitar darse cuenta a que paciente se le aplicó una u otra técnica. Por otro lado, en el caso de conducir estudios observacionales analíticos (estudios de cohortes y estudios de casos y controles<sup>23,24</sup>), dado que el investigador no controla las variables como ocurre en una situación experimental, no se puede asegurar que otros factores, no difieran entre uno y otro grupo en estudio; de este modo, la estimación del efecto del tratamiento pudiera estar sesgada por factores pronósticos; razón que hace que el nivel de evidencia de este tipo de estudios sea menor<sup>2,37</sup>. Ahora bien, es importante recalcar respecto de estos estudios, que la principal diferencia entre los estudios de cohorte y los estudios de casos y controles no reside en la secuencia temporal de la investigación, sino en los criterios de selección de las poblaciones sometidas a estudio; sin embargo, y dado que se parte de la existencia de un "evento de interés" o de una enfermedad, existe dificultad para asegurar una secuencia temporal lógica, donde la exposición antecede al efecto<sup>24</sup>. Ambos diseños tienen la desventaja de estar sujetos a varios sesgos, entre los que destacan los sesgos de selección, de entrevistador y de recuerdo, entre otros.

También podemos encontrar evaluación de tratamientos o PT realizados con estudios observacionales descriptivos; es decir, aquellos diseños en los que no existe grupo control o grupo de referencia con el cual comparar (algunos estudios de corte transversal, estudios correlacionales, series de casos y reporte de casos). Este tipo de estudios son, como se ha mencionado, los más comunes de encontrar en las revistas científicas<sup>3-6</sup>. Por su parte, los estudios de corte transversal nos dan una imagen instantánea de la coexistencia de la exposición y el efecto, y tienen las mismas limitaciones metodológicas del

estudio de casos y controles; además, tienen mayor dificultad para demostrar la secuencia temporal de causa y efecto<sup>38</sup>. Finalmente, las series y reporte de casos son de utilidad para describir los resultados observados en un paciente o un grupo de pacientes con un problema de salud similar; tomando en consideración que se trata de experiencias limitadas a las observaciones realizadas por un investigador o un grupo de ellos, desprovistas de un grupo de control o comparación; situación que asociada a una altísima probabilidad de todo tipo de sesgos, otorgan a este tipo de estudios un bajo nivel de evidencia en todas las clasificaciones en uso hoy<sup>37</sup>.

### **Aplicación de herramientas estadísticas y medidas de asociación**

No parece necesario a estas alturas del manuscrito entrar en detalles acerca de todas las herramientas estadísticas disponibles y acerca de aquellas que pueden utilizarse en artículos relacionados con tratamiento o PT.

No obstante ello, parece razonable recordar que existe un “hilo conductor” que surgiere comenzar siempre con la descripción de la muestra estudiada, es decir, aplicar la denominada estadística descriptiva, con cálculo de porcentajes, medidas de tendencia central (promedio, mediana, moda) y dispersión (rango, varianza y desviación estándar), determinación de prevalencia; estimación de IC 95%, curvas de supervivencia, etc; posteriormente aplicar los análisis bivariados, utilizando los conocidos  $\chi^2$  de Pearson y el exacto de Fisher para la comparación de los valores de frecuencia entre variables categóricas; pruebas paramétricas como t-Test o t de Student y ANOVA entre otros para la comparación de promedios; pruebas no-paramétricas para la comparación de variables de distribución asimétrica; y aplicar análisis multivariados, utilizando modelos de regresión lineal, logística u ordinal según sea el caso.

Todo lo anteriormente expuesto dependerá del tipo de diseño, las características de la población y de las variables con las que se encuentre trabajando el grupo de investigación.

No obstante ello, parece fundamental recalcar el concepto que un valor de  $p$  estadística o no estadísticamente significativo debe ser valorado en cada contexto, pues no es otra cosa que un valor que puede o no ser “estadísticamente significativo”; y no necesariamente se asocia a la multifactorial dinámica de la clínica. No es infrecuente observar en algunos artículos que un valor de  $p$  de 0,045 se considera “estadísticamente significativo”; lo que en rigor puede ser. Sin embargo, antes de asegurarlo

hay que valorar las características de la población, el tamaño de la muestra empleado para la conducción del estudio, qué herramientas estadísticas fueron utilizadas, etc.

Posterior a todo lo anteriormente expuesto, y si se trata de EC, estudios de cohortes y estudios de casos y controles; se ha de valorar la magnitud del efecto del tratamiento o PT en estudio respecto del estándar en uso o de otro; para lo cual, existen algunas herramientas con las que se puede comparar el riesgo en el grupo intervenido con el riesgo en el grupo control. Estas son las denominadas medidas de efecto (basadas en el cociente) y las medidas de impacto (basadas en la diferencia).

Las medidas de efecto son la estimación del “riesgo relativo” (RR) y del “odds ratio” (OR). La utilización de uno u otro dependerá del diseño de estudio con que se esté trabajando; de este modo, si se trata de un estudio de casos y controles, corresponderá aplicar OR; y si se trata de un EC o de un estudio de cohortes, corresponderá aplicar RR.

El OR no es otra cosa que el cociente entre la probabilidad de que ocurra el evento y la probabilidad de que no ocurra (odds); por ende, indica cuánto más probable es la ocurrencia del evento que su no ocurrencia. No tiene dimensiones, por lo que su rango va de 0 a infinito; y en resumen opera de la siguiente forma: cuando el OR es igual a 1, significa que no existe asociación; cuando el OR tiene un valor mayor a 1, significa que la asociación es positiva (es decir, que la presencia del factor se asocia a mayor ocurrencia del evento); y cuando el OR tiene un valor menor a 1, significa que la asociación es negativa (es decir, que la presencia del factor se asocia a menor ocurrencia del evento). Figuras 1 y 2.

Por otro lado, el riesgo expresa la probabilidad de un resultado adverso. Se expresa en unidades que van de 0 a 1 (es decir, sin riesgo a un riesgo de 100%). Requiere de un período de referencia y refleja la incidencia acumulada de una enfermedad o evento de interés en ese período de tiempo. Es así como nace el concepto de riesgo absoluto (RA) o incidencia que corresponde a una proporción que se puede definir como el número de personas que presentan el evento de interés en un momento dado de tiempo (eventos nuevos) sobre el número de personas en riesgo en ese momento. Entonces, la incidencia o RA siempre es calculada para un período de tiempo.

El RR es un cociente entre el riesgo en el grupo con el factor en estudio y el riesgo en el grupo de referencia. Es una relación (ratio) entre el riesgo de ocurrencia de cierto evento en el grupo expuesto o intervenido, comparado con el grupo control. No tiene dimensiones, por lo que su rango va de 0 a infinito; y en resumen opera de la siguiente forma:

	Casos	Controles
Expuestos	a	b
No expuestos	c	d

  

$$OR = \frac{a \times d}{b \times c}$$

**Figura 1.** Diagrama de función del cálculo de OR.

	Con morbilidad	Sin morbilidad
Con antecedentes	8	16
Sin antecedentes	25	153

  

$$OR = \frac{8 \times 153}{16 \times 25} = 3,06$$

**Figura 2.** Ejemplo de aplicación del cálculo de OR. Antecedentes de cirugía previa por hidatidosis hepática y desarrollo de morbilidad postoperatoria.

	Grupo estudio (N=32)	Grupo control (N=33)
Con ISO	7	8
Sin ISO	25	25

  

RA de ISO en grupo control =  $8 / 33 = 0,24$   
 RA de ISO en grupo estudio =  $7 / 32 = 0,22$   
 RR =  $0,22 / 0,24 = 0,91$   
 RR < 1. Es decir que la asociación es negativa  
 RAR =  $0,24 - 0,22 = 0,02$   
 RAR > 0. Es decir que la asociación es negativa  
 NNT =  $1 / 0,02 = 50$

**Figura 3.** Diagrama de función del cálculo de RR, RAR y NNT con un ejemplo referente a la efectividad de un tratamiento antibiótico acordado en pacientes con apendicitis aguda complicada respecto del tratamiento convencional o estándar.

cuando el RR es igual a 1, significa que no existe asociación; cuando el RR tiene un valor mayor a 1, significa que la asociación es positiva (es decir, que la presencia del factor en estudio se asocia a mayor ocurrencia del evento); y cuando el RR tiene un valor menor a 1, significa que la asociación es negativa (es decir, que la presencia del factor en estudio se asocia a menor ocurrencia del evento). Figura 3.

Las medidas de impacto por su parte son la “reducción absoluta de riesgo” (RAR), la “reducción relativa de riesgo” (RRR), el “número necesario a tratar” (NNT) y el “número necesario a dañar” (NND).

La RAR expresa qué tanto reduce el riesgo la intervención en estudio respecto de los sujetos que no la reciben. En otras palabras, corresponde a la

diferencia entre el riesgo en el grupo control y el riesgo en el grupo con el factor en estudio. No tiene dimensiones y su rango va de -1 a 1 porque se calcula restando el RR al valor 1 que representa al 100% de los sujetos. En resumen opera de la siguiente forma: cuando la RAR es igual a 0, significa que no existe asociación; cuando la RAR tiene un valor menor a 0, significa que la asociación es positiva (es decir, que la presencia del factor se asocia a mayor ocurrencia del evento); y cuando la RAR tiene un valor mayor a 0, significa que la asociación es negativa (es decir, que la presencia del factor se asocia a menor ocurrencia del evento).

La RRR, también denominada fracción atribuible o diferencia relativa de riesgo, es el cociente entre la disminución absoluta del riesgo y el riesgo del grupo control o lo que es lo mismo, diferencia entre el riesgo del grupo de prueba menos el riesgo del grupo control o de prueba menos el riesgo del grupo control o tratamiento estándar, dividido por el riesgo en el grupo control.  $RRR = RAR/Rc = (Rc-Re)/Rc$  tiene las mismas características que la RAR. Sin embargo, presenta una desventaja, y es que no diferencia los riesgos o beneficios muy grandes de los muy pequeños y no varía de acuerdo al tamaño de la muestra.

El NNT es un término introducido por Laupacis en 1988. Fue propuesto en el contexto de los ensayos clínicos para evaluar el impacto de un tratamiento. Se define como el número de individuos que hay que tratar con la terapia experimental para producir, o evitar, un evento adicional respecto a los que se producirían con el tratamiento control. Se calcula fácilmente, ya que es el inverso de la RAR, o  $1/RAR^{39}$ .

El NND representa lo contrario del NNT. Es decir, que un NNT negativo indica que el tratamiento tiene un efecto perjudicial (el tratamiento experimental consigue menos beneficio que el control o el

estándar), o que los efectos adversos inherentes al tratamiento son mayores en el grupo experimental. En otras palabras, el NNT representa el número necesario de personas a tratar para producir efecto en 1 de ellas; y el NND al número necesario de personas a tratar para producir algún daño en 1 de ellas. Cuanto más bajo sea el NNT, mayor será la magnitud del efecto de tratamiento en cuestión. Cuanto mayor es el NND, menor será el riesgo de dañar con el nuevo tratamiento o PT. Estos cálculos permiten evaluar no sólo la magnitud de los efectos sino también el costo-beneficio de la intervención. Si el NNT de un fármaco tiene un valor cercano al NND, las posibilidades de mejorar al paciente son similares a las posibilidades de provocarle algún daño; por lo que dicha droga tiene pocas posibilidades de ser útil en la práctica. Los EC bien realizados incluyen el NNT y el NND; o al menos incluyen los datos necesarios para efectuar los cálculos. Existen calculadoras online que permiten obtener fácilmente el NNT y el NND con sus respectivos IC 95% (<http://genesis.sefh.es/Enlaces/Calculadoras.htm>).

Todo artículo debe enunciar las herramientas estadísticas utilizadas en el proceso de análisis del estudio; desde las más simples a las más complejas. Y así como la mayor parte de los artículos mencionan herramientas estadísticas descriptivas y analíticas, tanto para realizar análisis bivariados como multivariados; es poco usual que se mencione la utilización de herramientas para valorar la magnitud del efecto del tratamiento o PT en estudio respecto del estándar en uso o de otro, con sus respectivos IC 95%. Si se trata de estudios comparativos, es exigible el reporte de este tipo de herramientas, o al menos que los autores publiquen las cifras con las que un lector pueda obtener los valores de las medidas de efecto y de impacto.

De este modo, las formas más apropiadas de representar los resultados en investigación clínica son los OR, el RA, el RR, la RAR, la RRR, el NNT y el NND<sup>39-41</sup>; y la significación estadística no es nada más que eso, “la significación estadística”, que en ocasiones puede ser positiva y clínicamente irrelevante, o negativa, sin que eso signifique necesariamente que no hay diferencias reales entre las variables en estudio.

### **Sistemas de puntuación y listas de comprobación**

A continuación se mencionan iniciativas de distintos grupos de estudio de metodología de la investigación, los que han aportado diferentes herramientas que ayudan en la valoración general y específica de la calidad metodológica de los artículos.

CONSORT: “Consolidated Standards of Reporting Trials”. Fue desarrollado para guiar a los autores a mejorar la calidad de la publicación de EC controlados con asignación aleatoria. Es una lista de comprobación que consta de 5 dominios (Título y resumen, introducción, métodos, resultados y discusión), que incluyen 22 ítems, en los que se evalúa la descripción de una serie de variables inherentes a un EC. Entre los ítems, solicita de los autores la creación de un diagrama de flujo en el que se describa los pasos de los participantes en el estudio, desde la selección y reclutamiento, la distribución del tratamiento, el seguimiento y el análisis. Es una de las herramientas más utilizadas y se encuentra en permanente actualización<sup>9,10</sup>.

QUORUM: “Quality of Reporting of Meta-analyses”. Fue desarrollado para guiar a los autores a mejorar la calidad de la publicación de un meta-análisis. Es una lista de comprobación que consta de 6 dominios (resumen, introducción, métodos, resultados y discusión de un meta-análisis), que está organizada en 21 categorías y subcategorías relativas a; búsquedas, selección de artículos primarios, evaluación de validez de los artículos, extracción de datos, características de los estudios, síntesis de los datos cuantitativos metodológicos, etc. También se requiere de un diagrama de flujo que proporcione información respecto de los EC incluidos y excluidos, y las razones para la exclusión de estos<sup>11</sup>.

STROBE: “Strengthening the Reporting of Observational Studies in Epidemiology”. Fue desarrollado para guiar a los autores a mejorar la calidad de la publicación de estudios observacionales. Esta declaración, consta de 5 dominios (Título y resumen, introducción, métodos, resultados y discusión), que incluyen 22 ítems; 18 de los cuales son de aplicación general para estudios de cohortes, de casos y controles, y de corte transversal; y 4 que son específicos para cada uno de los tres diseños antes mencionados. También requiere por parte de los autores de la generación de un diagrama de flujo<sup>12</sup>.

TREND: Fue desarrollado para guiar a los autores a mejorar la calidad de la publicación de estudios que utilizarán diseños sin asignación aleatoria. Esta declaración, consta de 4 dominios (Título, resumen e introducción, métodos, resultados (incluye un diagrama de flujo de los participantes) y discusión; los que incluyen 22 ítems. Consiste en evaluar EC sin asignación aleatoria; y sus directrices hacen hincapié en la presentación de las teorías utilizadas, la descripción de la intervención, las condiciones de comparación, el diseño de investigación utilizado y los métodos de ajuste para posibles sesgos en los estudios que utilizan diseños sin asignación aleatoria<sup>13</sup>.

MINCIR: “Metodología de Investigación en Cirugía”. Fue desarrollado para valorar la calidad

metodológica de estudios de distintos diseños en relación a artículos de tratamiento o PT. Está compuesto por 3 dominios: el primero, relacionado con el diseño del estudio; el segundo, con el tamaño de la población estudiada; y el tercero, relacionado con la descripción de la metodología empleada en el estudio (mención de los objetivos, justificación del diseño, criterios de elegibilidad de la muestra y justificación de la misma). De esta forma, se genera una puntuación final que puede fluctuar entre 6 y 36 puntos, asignando 6 puntos al estudio de menor calidad metodológica y 36 puntos al de mejor calidad metodológica<sup>4,42,43</sup>.

MINORS: "Methodological index for non-randomized studies". Fue desarrollado para guiar a los autores a mejorar la calidad de la publicación de estudios sin asignación aleatoria en el ámbito quirúrgico; sean comparativos o no. Contiene 12 ítems, los primeros ocho para los estudios no comparativos, los restantes ítems para los estudios comparativos<sup>44</sup>.

En resumen quien escribe un manuscrito relacionado con tratamiento o PT, o quien lee un artículo de este tipo debe exigir al menos un objetivo claro, preciso y conciso respecto del escenario de la investigación que se realizó; la mención explícita del diseño empleado con los respectivos detalles metodológicos inherentes a este; y la mención y ejecución de herramientas estadísticas y medidas de asociación, o al menos los números necesarios para poder calcular estos valores.

## Referencias

1. Manterola C. Medicina Basada en la Evidencia o Medicina Basada en Pruebas. Generalidades acerca de su aplicación en la práctica clínica cotidiana. *Rev Med CLC* 2009; 20: 125-130.
2. Oxford Centre for Evidence-based Medicine - Levels of Evidence (March 2009). <http://www.cebm.net/index.aspx?o=1025>. Consultado el 21 de marzo de 2010.
3. 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: 500-507.
4. Manterola C, Pineda V, Vial M, Losada H; MINCIR Group. What is the methodologic quality of human therapy studies in ISI surgical publications? *Ann Surg* 2006; 244: 827-832.
5. 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.
6. Manterola C, Grande L. Methodological quality of articles on therapeutic procedures published in Cirugía Española. Evaluation of the period 2005-2008. *Cir Esp* 2010 Jan 13. [Epub ahead of print].
7. Manterola C, Zavando D; Grupo MINCIR. Cómo interpretar los "Niveles de Evidencia" en los diferentes escenarios clínicos. *Rev Chil Cir* 2009; 61: 582-595.
8. Manterola C, Pineda V, Vial M, Grande L. How should the results of a research study be presented? II. The manuscript and the publication process. *Cir Esp* 2007; 81: 70-77.
9. Moher D, Schulz KF, Altman D; CONSORT Group (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001; 285: 1987-1991.
10. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al (CONSORT group). Pragmatic Trials in Healthcare (Practihc) group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008; 337: a2390. doi: 10.1136/bmj.a2390.
11. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 1999; 354 (9193): 1896-1900.
12. Vandembroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al (STROBE initiative). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med* 2007; 147: W163-194.
13. Des Jarlais DC, Lyles C, Crepaz N (TREND Group). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health* 2004; 94: 361-366.
14. 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: 404-411.
15. Manterola C, Pineda V, Vial M. Resección abierta vs Laparoscópica en cáncer de colon no complicado. Revisión sistemática. *Cir Esp* 2005; 78: 17-22.
16. Yang H, Watson DI, Lally CJ, Devitt PG, Game PA, Jamieson GG. Randomized trial of division versus non-division of the short gastric vessels during laparoscopic Nissen fundoplication: 10-year outcomes. *Ann Surg* 2008; 247: 38-42.
17. Rossi M, Barreca M, De Bortoli N, Renzi C, Santi S, Gennai A, et al. Efficacy of Nissen fundoplication versus medical therapy in the regression of low-grade dysplasia in patients with Barrett esophagus: a prospective study. *Ann Surg* 2006; 243: 58-63.
18. Frazzoni M, Manno M, De Micheli E, Savarino V. Intra-oesophageal acid suppression in complicated gastro-oesophageal reflux disease: esomeprazole versus lansoprazole. *Dig Liver Dis* 2006; 38: 85-90.

19. Koksall N, Muftuoglu T, Gunerhan Y, Uzun MA, Kurt R. Management of intrabiliary ruptured hydatid disease of the liver. *Hepatogastroenterology* 2001; 48: 1094-1096.
20. Hamamci EO, Besim H, Sonisik M, Korkmaz A. Occult intrabiliary rupture of hydatid cysts in the liver. *World J Surg* 2005; 29: 224-226.
21. <http://www.fisterra.com/guias2/fmc/preguntas.asp>. Accesada el 21 de marzo de 2010.
22. Liu P, Yang JM, Niu WY, Kan T, Xie F, Li DQ, et al. Prognostic factors in the surgical treatment of caudate lobe hepatocellular carcinoma. *World J Gastroenterol* 2010; 16: 1123-1128.
23. Manterola C, Muñoz S, Bustos L. Estrategias de investigación. Diseños observacionales. 3ª parte. Estudio de cohortes. *Rev Chil Cir* 2001; 53: 410-414.
24. Manterola C. Estudios observacionales. Los diseños utilizados con mayor frecuencia en investigación clínica. *Rev Med CLC* 2009; 20: 539-548.
25. Manterola C, Bustos L. Estrategias de investigación. Diseños experimentales. Ensayo clínico. *Rev Chil Cir* 2001; 53: 498-503.
26. Manterola C, Pineda V, Vial M. Acerca de los ensayos clínicos en el ámbito quirúrgico. Reflexiones, dificultades y desafíos. *Rev Chil Cir* 2009; 61: 482-489.
27. Manterola C, Grande L. Methodological quality of articles on therapeutic procedures published in *Cirugía Española*. Evaluation of the period 2005-2008. *Cir Esp* 2010 Jan 13. [Epub ahead of print].
28. Hennekens CH, Buring JE. *Epidemiology in Medicine*. Boston: Little Brown and Company; 1987.
29. Kelsey JL, Thompson WD, Evans AS. *Methods in Observational Epidemiology*. New York: Oxford University Press; 1986.
30. Feinstein AR. An outline of cause-effect evaluations En: *Clinical epidemiology the architecture of clinical research*. W.B. Saunders Company; 1985. pp. 39-52.
31. Manterola C, Pineda V, Grupo MINCIR. El valor de “p” y la “significación estadística”. Aspectos generales y su valor en la práctica clínica. *Rev Chil Cir* 2008; 60: 86-89.
32. Riegelman RK. Studying a study and testing a test: how to read the medical evidence. 4th ed. Philadelphia, Pa.: Lippincott Williams & Wilkins; 2000.
33. Dawson-Saunders B, Trapp, RG. Obtención de inferencias a partir de datos: Intervalos de confianza y pruebas de hipótesis. En: *Bioestadística médica*. 2ª ed. México. El Manual Moderno; 1997. pp. 99-118.
34. Castellanos-Olivares A, Vásquez-Márquez I, Wachter-Rodarte N. ¿Qué es un sesgo? En: Ramiro HM, Saita-Kamino O, editores. *Temas de medicina interna, epidemiología clínica*. México. Interamericana-McGraw-Hill; 1994. pp. 237-247.
35. Fletcher RH, Fletcher SW, Wagner EH. *Epidemiología clínica. Aspectos fundamentales*. 2ª ed. México: Masson Williams (Wilkins, España, SA); 1998. pp. 7-10).
36. Manterola C. Revisión sistemática de la literatura síntesis de la evidencia. *Rev Med Clin Condes* 2009; 20: 897-903.
37. Manterola C, Zavando D, Grupo MINCIR. Cómo interpretar los “Niveles de Evidencia” en los diferentes escenarios clínicos. *Rev Chil Cir* 2009; 61: 582-595.
38. Hernández B, Velasco-Mondragón HE. Encuestas transversales *Salud Pública Mex* 2000; 42:447-455.
39. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988; 318: 1728-1733.
40. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ* 1995; 310: 452-454.
41. Sackett DL, Richardson WS, Rosenberg W, Hynes RB. *Evidence-based medicine: how to practice and teach EBM*. Churchill-livingstone; London, 2nd ed. 2000.
42. Manterola C, Pineda V, Vial M, Losada H. Revisión crítica de la literatura para artículos de terapia. *Rev Chil Cir* 2004; 56: 604-609.
43. Manterola C, Vial M, Pineda V, Sanhueza A. Systematic Review of Literature with Different Types of Designs. *Int J Morphol* 2009; 27: 1179-1186.
44. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (MINORS): development and validation of a new instrument. *ANZ J Surg* 2003; 73: 712-716.