

DOCUMENTOS

Revisión sistemática de la literatura. Propuesta metodológica para su realización

Drs. CARLOS MANTEROLA D, VIVIANA PINEDA N, MANUEL VIAL G,
HÉCTOR LOSADA M, SERGIO MUÑOZ N

Departamento de Cirugía, CIGES*, Facultad de Medicina, Universidad de La Frontera

INTRODUCCIÓN

¿Cómo elegir una terapia basada en la evidencia? Probablemente es una pregunta que con frecuencia cada vez mayor nos hacemos en la práctica cotidiana de la medicina. Para responder a ésta, es imprescindible desarrollar el esquema de la Figura 1, en el que el paso número uno es definir el problema clínico e inmediatamente transformarlo en una pregunta contestable, de esta forma podremos dar respuesta a algo (es obvio que quién nada se pregunta nada se responderá). A continuación, y con una pregunta bien definida podremos buscar una respuesta, para lo cual deberemos buscar en la fuente apropiada el conocimiento que necesitamos para responder a nuestra inquietud, luego tendremos que tener, o de lo contrario diseñar, una estrategia de búsqueda y finalmente resumir la evidencia encontrada según su nivel, calidad metodológica, validez y confiabilidad. Este proceso lo podríamos entonces resumir en una sola oración: "buscar con la máxima eficiencia la mejor evidencia disponible para responder a una pregunta de investigación".

Pero, ¿qué es el nivel de evidencia? No es otra cosa que una escala con la que se intenta "calificar" la evidencia disponible, para lo cual se jerarquizan los diseños de investigación existentes en clínica según la solidez del aporte que pueden generar en 4 ámbitos del conocimiento clínico: terapia, prevención, etiología y daño; pronóstico, diagnóstico, y análisis económicos (Tabla 1). Como toda escala,

ésta está constituida por valores numéricos, cuya significación podría traducirse en "excelente", "buena", "regular", "mala", "muy mala", etcétera, pero que nos provee de una herramienta útil para referirnos a la evidencia que apoye nuestro accionar. Un ejemplo es la escala de niveles de evidencia propuesta por Sackett,¹ en la que los niveles se clasifican de 1 a 5, siendo el Nivel 1 el que aporta la mejor evidencia; y el Nivel 5, el que aporta la evidencia "menos buena" o "la más mala" (según la que usted prefiera).

Y. ¿Qué es validez y confiabilidad? La validez se puede definir como el grado con que los resultados de una medición se corresponden con el estado de veracidad del fenómeno medido. Son sinónimos exactitud y precisión; y la confiabilidad o reproducibilidad significa que las mediciones realizadas en ocasiones diferentes, o por observadores diferentes, o por pruebas paralelas, determinarán los mismos resultados.^{2,3}

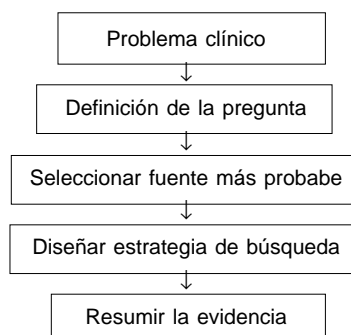


Figura 1. Del problema clínico cotidiano al resumen de la evidencia.

* Capacitación, Investigación y Gestión para la salud basada en evidencia.

Tabla 1
NIVELES DE EVIDENCIA PROPUESTOS POR SACKETT PARA ESTUDIOS SOBRE TERAPIA, PREVENCIÓN, ETIOLOGÍA Y DAÑO

Recomendación	Nivel de evidencia	Tipo de estudios sobre los que se basa	
A	Nivel 1	1a	Meta-análisis
		1b	Ensayo clínico aleatorio controlado individual con intervalo de confianza estrecho
B	Nivel 2	2a	Revisión sistemática de estudios de cohortes
		2b	Estudios de cohortes individual. Ensayo clínico aleatorio de baja calidad
		2c	Investigación de "desenlaces"
C	Nivel 3	3a	Revisión sistemática de estudios de casos y controles
		3b	Estudio individual de casos y controles
D	Nivel 4	Series de casos. Estudios de cohortes y de casos y controles de mala calidad	
D	Nivel 5	Opinión de expertos sin evaluación crítica explícita	

Después de recordar estos conceptos volvamos a la pregunta original. ¿Cómo elegir una terapia basada en la evidencia?, más aún, ¿cómo buscar la mejor evidencia para una terapia determinada de una manera eficiente? Si miramos la Tabla 1, podemos entender que con un nivel de evidencia 1 tanto realizando un ensayo clínico aleatorio individual (ECA) como una revisión sistemática de la literatura (RSL) que utilice ECAs de buena calidad (también conocidos como meta-análisis) podremos responder a preguntas sobre terapia con un buen nivel de evidencia y en forma eficiente. Sin embargo, en el mundo quirúrgico la conducción de ECAs de doble ciego es muy difícil, pues si recordamos que "los potenciales ciegos" son el que administra la terapia, el que la recibe y el que mide los resultados de ésta, podemos comprender que para tener un doble ciego en una investigación de terapia que involucre una herida o cicatriz sólo podríamos comparar técnicas quirúrgicas que se realicen en forma similar y por un acceso más o menos común, de lo contrario para que el que reciba o mida el efecto de la terapia estén ciegos tendríamos que realizarle a los pacientes del grupo de control una herida similar a los del grupo en estudio, hecho que es inaceptable desde el punto de vista ético (esto sin considerar los aspectos económicos involucrados en la investigación, la dificultad técnica que involucra este tipo de diseño, etcétera). Por lo tanto, la otra opción es intentar responder preguntas relacionadas con terapia a través de "estudio de estudios", es decir mediante la práctica de meta-análisis o de RSL, que no son otra cosa que análisis cualitativos y cuantitativos de la literatura existente. Al respecto, hay que recordar que un meta-análisis tiene la ventaja de ser un diseño con nivel de evidencia 1, pero

que consiste en el análisis comparativo de ECAs, en el que se permite el estudio de estudios que examinan la misma pregunta y el efecto de una misma terapia, hecho que no es fácil de encontrar en el mundo quirúrgico; por lo tanto, una opción intermedia es la realización de revisiones sistemáticas de la literatura que pueden aportar un nivel de evidencia 2 ó 3, pues, de esta forma se puede trabajar con diseños de investigación variados (no sólo con meta-análisis), se pueden examinar distintas preguntas u objetivos, y estudiar el efecto de distintas terapias, razones por las cuales es a nuestro modo de ver una alternativa apropiada para ayudarnos a buscar la mejor evidencia para una terapia determinada de una manera eficiente (Figura 2).

METODOLOGÍA

Una vez que se ha tomado la decisión de realizar una RSL, se ha de precisar la población de estudios primarios con la que se ha de trabajar, por lo que debemos precisar los criterios de inclusión y exclusión; por ejemplo, podemos incluir estudios en los que se reporten resultados de la aplicación de una terapia a individuos humanos y adultos, definiremos cuáles serán las variables de respuesta a estudiar, el tipo de diseños de investigación que aceptaremos, los idiomas originales de las publicaciones que analizaremos y el período de tiempo de la revisión (año de inicio y de término de la búsqueda de información), por ejemplo, idiomas inglés y español, entre 1990 y 2002. Por otra parte, podremos excluir artículos de temática no antigente al tema en estudio, aquellos cuyo resumen no se encuentre disponible en la base de datos en la que

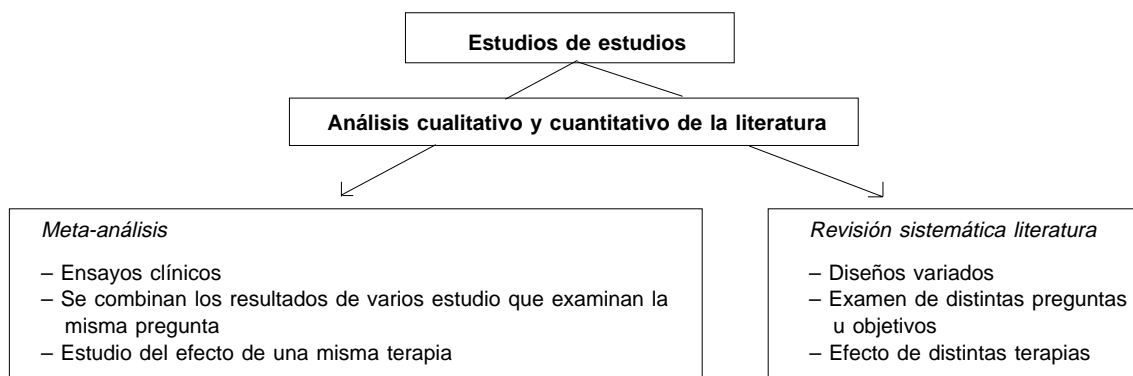


Figura 2. ¿Cómo buscar la mejor evidencia para una terapia en forma eficiente?

desarrollamos la búsqueda, artículos de revisión, cartas al Editor, otras revisiones sistemáticas de la literatura, etc.

A continuación deberemos precisar cuál será nuestra estrategia de búsqueda de los estudios primarios (aquellos con los que vamos a trabajar). Para ello es necesario haber precisado la población blanco y los criterios de inclusión. Así, si nos interesa estudiar artículos de habla hispana y en especial del mundo latinoamericano tendremos necesariamente que acceder a la base de datos LILACs, utilizando “palabras libres”; pero, si lo que nos interesa se encuentra preferentemente en el mundo anglosajón tendremos que utilizar índices como MedLine recurriendo a los denominados términos MeSH (Medical Headings) y también a “palabras libres”, de este modo nos aseguramos de encontrar un gran porcentaje de la literatura publicada. Una forma adicional de asegurar la menor pérdida posible de información es mediante la búsqueda manual de revistas de interés particular al tema en cuestión.

Y, ¿cómo hacer la revisión y ulterior comparación de los datos obtenidos a partir de los artículos primarios? Primero se ha de conformar un equipo de trabajo, pues no es planteable realizar este tipo de estudios en forma aislada, ya que se sabe que esto es una fuente importante de sesgo en este tipo de investigación. En segundo lugar, se considera deseable que el análisis de los artículos se realice en forma independiente al menos por 2 investigadores, que en casos de discrepancias se recurra a la opinión de un tercer investigador y que de persistir éstas, se llegue a una decisión por consenso. En tercer lugar, tenemos que pensar en cómo realizar un análisis de la calidad metodológica de los estudios (que suele ser muy diversa) y de qué forma comparar resultados difíciles de comparar debido al tipo de poblaciones estudiadas, tamaños de la

muestra, variables medidas, etcétera de tal forma que los resultados publicados en los estudios primarios queden adecuadamente representados independientemente de la calidad metodológica de cada estudio, así que a un ECA se le asigne el peso que tiene en relación con una serie de casos; que a un estudio con un “n” de 2000 sujetos tratados se le conceda el peso que corresponde al compararlo con un estudio con un “n” de 25 casos; que a un estudio que tiene objetivos claros y concretos, que define criterios de selección, justifica la muestra empleada, etcétera, se le valore en forma apropiada en relación con un estudio que carece de objetivos, o que no menciona criterios de selección de los pacientes a tratar, o no justifica el tamaño de la muestra empleado.

Para intentar salvar estas situaciones, que sin lugar a duda son difíciles de zanjar, y que requieren de algo más que imaginación, nuestro grupo de trabajo ha planeado a modo de propuesta de estudio 2 vertientes. La primera es la construcción de un score de calidad metodológica de los estudios primarios, y la segunda el “cálculo de promedios ponderados”.

El score al que se hace mención, contiene 3 ítems; el primero relacionado con el tipo de diseño del estudio, el segundo con el tamaño de la población estudiada ajustado por la justificación o no de este tamaño de muestra, y el tercero relacionado con la metodología empleada en el estudio en cuestión (objetivos, justificación del diseño, criterios de selección de la muestra y justificación de ésta). De esta forma, se genera una escala que es la sumatoria de los ítems 1, 2 y 3, cuyo puntaje total puede fluctuar entre 6 y 36 puntos, asignando 6 puntos al estudio de menor calidad metodológica y 36 puntos al de mejor calidad metodológica (Tabla 2).

El cálculo de promedios ponderados no es otra

Tabla 2
ESCORE DE CALIDAD METODOLÓGICA DISEÑADO PARA LA REVISIÓN SISTEMÁTICA DE LOS ESTUDIOS PRIMARIOS

<i>Item 1. Diseño del estudio</i>	<i>Puntaje asignado</i>
Ensayo clínico multicéntrico	12
Ensayo clínico aleatorio con doble ciego	9
Ensayo clínico (sin ciego o ciego simple, sin aleatorización)	6
Cohorte concurrente	4
Casos y controles y estudios de cohorte histórica	3
Corte transversal	3
Serie de casos	1
<i>Item 2. Población estudiada x factor de justificación</i> (FJ= 2 justifica la muestra y 1 no justifica)	<i>Puntaje asignado</i>
≥ 201	6 ó 12
151-200	5 ó 10
101-150	4 ó 8
61-100	3 ó 6
31-60	2 ó 4
≤ 30	1 ó 2
<i>Item 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
– Sólo 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
Escore final	<i>Puntaje total</i>
ítem 1 + (ítem 2 x factor justificación) + ítem 3	6 a 36

cosa, que el producto de la sumatoria de las variables en estudio por el escore de calidad metodológica del artículo que la originó, dividido por la sumatoria de los escores de calidad metodológica de todos los artículos en estudio (Figura 3). Con esta idea, pretendemos que aquellos artículos de buena calidad metodológica, con un diseño de buen nivel de evidencia tengan una representación justa a la hora de comparar resultados frente a un artículo de menor calidad metodológica, con un diseño de bajo nivel de evidencia, aun cuando este último tenga un mayor número de sujetos tratados. Esta idea se puede entender mejor al observar la Tabla 3, que corresponde a un extracto de una planilla Excel en la que se presenta a modo de

ejemplo el tratamiento de una variable en un estudio cualquiera.

En la primera columna se encuentra el nombre del autor principal, en la segunda el número de sujetos tratados, en la tercera el año de la publica-

$$PP_{GT} = \frac{\sum Xi \cdot ei}{\sum ei}$$

Xi: Valor variable en el estudio i (todas las variables).
 ei: Escore obtenido por el estudio i.
 Σei: Sumatoria de los escores de todos los estudios

Figura 3. Cálculo de promedios ponderados para las variables estudiadas de los grupos de terapia.

Tabla 3
EJEMPLO DEL CÁLCULO DE PROMEDIOS PONDERADOS PARA LA VARIABLE MORBILIDAD EN UNA RSL EN UN GRUPO DE TERAPIA CUALQUIERA

Autor	n	Año	Morbilidad (%)	Morbilidad*e	Escore (ptos)
Capizzi	80	2002	3,8	30,4	8
De Wit	25	1999	28	476,0	17
Morino	15	1997	60	420,0	7
Sauri	207	2000	12	156,0	13
Rubenstein	63	2002	28	336,0	12
Victorzon	60	2002	12	132,0	11
Favretti	830	2002	25,3	354,2	14
Baltasar 1	27	2002	33,3	199,8	6
Matthews	48	2000	12,5	125,0	10
Nguyen	79	2001	27,8*	556,0	20*
Schauer	275	2000	30,5	457,5	15
Rutledge	1274	2001	7	77,0	11
Baltasar 2	16	2002	43,8**	262,8	6**
Paiva	40	2002	25	175,0	7
Magnusson	154	2002	23,4	257,4	11
Ponderación				23,9*	168*

*Verde; **Rojo.

ción, en la cuarta la morbilidad reportada por el artículo primario, en la quinta columna el producto entre la variable morbilidad y el escore de calidad metodológica del artículo primario y en la última columna el escore de cada artículo en estudio. De esta forma se puede constatar la gran dispersión de morbilidad reportada por los diferentes artículos. Así, se puede apreciar en rojo que un artículo de escore de calidad muy bajo presenta una morbilidad alta, sin embargo en azul se observa como esta misma variable es mucho más baja en un artículo de mejor calidad metodológica. De esta forma, en la última línea se puede apreciar en la quinta columna y en verde, el promedio ponderado de la variable morbilidad (23,9%) para todo este grupo de artículos primarios, y que parece bastante adecuada al observar la morbilidad reportada por todos los artículos en estudio. Al lado de esta cifra, se aprecia también en verde 168, que es la sumatoria de los escores de todos los artículos analizados.

Con base en esta metodología, intentaremos responder en los próximos números de esta revista a las siguientes preguntas:

1. ¿Cuál es la mejor terapia para los pacientes

con coledocolitiasis con vesícula *in situ*, comparando procedimientos endoscópicos y quirúrgicos, laparoscópicos y abiertos?

2. ¿Cuál es la mejor terapia para los pacientes con esófago de Barrett al comparar procedimientos endoscópicos y quirúrgicos, laparoscópicos y abiertos?

3. ¿Cuál es la mejor opción quirúrgica para el tratamiento de pacientes con cáncer de colon no complicado, al comparar técnicas laparoscópicas y abiertas?

4. ¿Cuál es la mejor opción quirúrgica para el tratamiento de pacientes con obesidad mórbida, considerando técnicas abiertas y laparoscópicas?

BIBLIOGRAFÍA

1. Sackett DL, Wennberg JE: Choosing the best research design for each question. *BMJ* 1997; 315: 1636-40.
2. Streiner DL, Norman GR: Health measurement scales. A practical guide to their development and use. Oxford University Press. New York, United States, 1995; 1-3
3. Manterola C: El proceso de medición con variables cualitativas y su aplicación en cirugía. *Rev Chil Cir* 2002; 54: 307-15.