

Del Editor

Medwave 2012 Sep;12(8):e5481 doi: 10.5867/medwave.2012.08.5481

A propósito de las revisiones sistemáticas: dando pasos para una mejor interpretación de la evidencia

On systematic reviews: taking steps to improve evidence interpretation

Autor: Miguel Araujo Alonso⁽¹⁾

Filiación: ⁽¹⁾Hospital de Los Andes, Chile; Medwave, Santiago, Chile

E-mail: maraujo@medwave.cl

Citación: Araujo M. On systematic reviews: taking steps to improve evidence interpretation. Medwave 2012 Sep;12(8):e5481 doi: 10.5867/medwave.2012.08.5481

Fecha de envío: 29/8/2012

Fecha de aceptación: 30/8/2012

Fecha de publicación: 1/9/2012

Correspondencia a:

⁽¹⁾Villaseca 21, oficina 702, Ñuñoa, Santiago de Chile

Las revisiones sistemáticas (RS) se mantienen activas, y ello es favorable para la comunidad profesional y científica. A propósito de la inclusión en esta edición de un número importante de estos estudios, hemos querido incentivar la reflexión y la puesta al día en nuestros lectores, abordando dos aspectos que vale la pena conocer cuando se trata de RS.

Análisis secundarios y sus aplicaciones

Con la ampliación del campo de las RS a las intervenciones "complejas" –por ejemplo, programas comunitarios multicomponente-, que son de interés para adoptar decisiones de salud pública, y la constancia de que muchas veces las RS aportan resultados generales sobre su efectividad pero no dan cuenta de preguntas más específicas, por ejemplo sobre la forma óptima de entregar la intervención, o sobre quiénes se podrían beneficiar mejor de ellas, ha surgido la demanda por identificar atributos en algunas intervenciones, características de los participantes o elementos del entorno, que pudieran modificar su efecto, a través de métodos de análisis secundarios en el contexto de una RS o metaanálisis. Cuando una intervención es difícil de reproducir fielmente como en estos casos, se puede asumir que son los factores que influyen sobre el éxito de la intervención los que hay que identificar y aplicar al diseñar intervenciones localmente.

Suponga que desea conocer la duración óptima de una intervención de salud pública y dispone de varios estudios en los que ésta se aplicó durante períodos distintos (por ejemplo, en algunos estudios duró un mes, en otros tres meses, y en un tercer grupo se prolongó por seis o más meses). O que desea analizar por separado el efecto según sexo, incluyendo varios estudios a la vez, cada uno de los cuales tiene subgrupos de mujeres y hombres. Comparar los resultados entre los estudios o entre grupos

de estudios, con el fin de asociar la magnitud del efecto con alguna característica secundaria, puede conducir a conclusiones erróneas, porque además del factor evaluado los estudios pueden diferir en muchos otros aspectos, incluso ni siquiera ser cuantificables. En rigor, las comparaciones realmente confiables son aquellas realizadas en forma directa, "cara a cara", entre las dos ramas primarias al interior de un mismo ensayo. Por ejemplo, un estudio que compara dos grupos que reciben la misma intervención con distinta intensidad. O un ensayo que incluye un análisis de subgrupos definido a priori y con un análisis estadístico robusto. No obstante, este tipo de comparaciones es escaso, lo que obliga a intentar analizar interacciones a través de comparaciones indirectas, entre estudios.

Al problema tampoco están ajenos otros tipos de intervenciones. Por ejemplo, una de las quejas de muchos es la falta de evidencia sobre la efectividad relativa de fármacos que son alternativos, sobre los que no hay estudios que aporten comparaciones directas. Y también ocurre lo mismo cuando se trata de comparar métodos diagnósticos entre sí.

Para resolver este tipo de cuestiones se recurre, entre otros, a los métodos de análisis estratificado, por ejemplo, comparando el resultado combinado de los estudios en los que la intervención se aplicó de cierta forma, con el resultado combinado de otro grupo de estudios en que se aplicó de manera distinta. Con las conclusiones de estas comparaciones se debe ser más bien cauto, pero en muchos casos ayudan a tomar decisiones, por ejemplo, cuando hay una evidente diferencia en la magnitud del efecto entre las modalidades comparadas, o cuando el comportamiento del conjunto de los estudios muestra claras tendencias o patrones. El análisis de subgrupos combinando los resultados de varios estudios es otra posibilidad que ofrecen los metaanálisis, ya sea en forma

univariada, o cada vez más, utilizando las técnicas de meta-regresión, que permiten ajustar varias características de los estudios simultáneamente. En ocasiones también, los métodos se plantean desde la perspectiva de los análisis de sensibilidad, que al revés, no pretenden encontrar fuentes de variación, sino más bien demostrar la estabilidad o consistencia de los efectos bajo distintos escenarios.

Estas técnicas están sirviendo, si no para agregar validez a las RS, para obtener mejor provecho de la información que proveen los estudios para apoyar la toma de decisiones. Las comparaciones entre estudios ayudan a comprender mejor los efectos de las intervenciones, y los lectores deben intentar conocer sus limitaciones y potencialidades. En nuestra revista dedicaremos una de nuestras próximas series de artículos metodológicos a estos aspectos.

Evidencia empírica de los sesgos

Una variante interesante de los análisis secundarios de las RS es la que trata de relacionar la calidad de los estudios con sus resultados. A través de estos análisis, al comparar los resultados entre estudios de distinto diseño o calidad, se ha acumulado cada vez más evidencia empírica de que los estudios de menor calidad tienden a sobreestimar el efecto de las intervenciones (así como la importancia de los factores de riesgo), lo que traducido en políticas públicas podría llevar a malgastar recursos en medidas poco efectivas o simplemente inútiles¹. En oposición, los estudios de mejor calidad arrojan con mayor frecuencia resultados negativos (NS). Los ejemplos sobre este tipo de sesgo son múltiples y lo esperable es que la mayoría de los ámbitos de investigación en salud se vean afectados por él. Tópicos tan dispares como el riesgo de cáncer gástrico asociado al consumo de carne procesada² y las medidas de prevención de accidentes en motocicleta³ han mostrado resultados similares. Los mejores diseños (ejemplo: ensayos aleatorizados vs estudios cuasiexperimentales, cohortes vs estudios caso-control), y dentro de un mismo diseño, los estudios con mejores atributos metodológicos (ejemplo: doble ciego vs estudio abierto, cohorte sin ajuste vs cohorte ajustada), consistentemente exhiben la misma tendencia. Inclusive analizando estudios observacionales que tratan de un mismo tema se han observado relaciones dosis-respuesta, donde a mayor número de covariables utilizadas para ajustar los resultados, menor es el tamaño de efecto observado.

Cuando se interpretan resultados, por lo tanto, no es indiferente qué tipo y calidad de estudios estamos analizando. Esto refuerza la necesidad de que, frente a

distintas variedades de estudio, se privilegie los resultados de los mejores diseños, porque los estudios de inferior calidad posiblemente muestren efectos superiores en promedio a los reales.

Por lo mismo, ha sido materia de debate en las revisiones que admiten no sólo ensayos aleatorizados “dónde detenerse” en cuanto al diseño de los estudios seleccionados. La mayoría da por sentado que limitar la revisión a los ensayos aleatorizados en muchos casos puede ser demasiado restrictivo, pero el límite inferior no ha sido fijado. Por otro lado, también se ha establecido que en gran parte de los casos los estudios observacionales coinciden al menos en la dirección de los efectos, con los resultados de los ensayos clínicos⁴. La cuestión no es menor. Algunos han denominado “inverse evidence law” al fenómeno de que la evidencia sobre las intervenciones que podrían tener mayor impacto potencial a escala de salud pública sea la menos abundante, mientras lo que se hace visible como evidencia de mayor peso sean estudios de bajo impacto masivo o poco generalizables, pero “metodológicamente correctos”.

Una crítica frecuente a las revisiones sistemáticas Cochrane, es que en demasiadas oportunidades la conclusión es simplemente que “no existe evidencia suficiente...”. Sin desconocerlo, también debemos reconocer que las RS han aportado luces sobre el verdadero potencial de sesgo de los estudios, nos enseñan a ser más cautos frente a ello, y que, por otra parte, se ha avanzado en el desarrollo de métodos de análisis que enriquecen la interpretación de sus resultados.

Referencias

1. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*. 352(9128):609-13, 1998 Aug 22. ↑ | [CrossRef](#) | [PubMed](#) |
2. Larsson SC, Orsini N, Wolk A. Processed meat consumption and stomach cancer risk: a meta-analysis. *J Natl Cancer Inst*. 2006 Aug 2;98(15):1078-87. ↑ | [CrossRef](#) | [PubMed](#) |
3. Elvik R. Evaluations of road accident blackspot treatment: a case of the Iron Law of Evaluation Studies? *Accid Anal Prev*. 1997 Mar;29(2):191-9. ↑ | [CrossRef](#) | [PubMed](#) |
4. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000 Jun 22;342(25):1878-86. ↑ | [CrossRef](#) | [PubMed](#) |



Esta obra de Medwave está bajo una licencia Creative Commons Atribución-No Comercial 3.0 Unported. Esta licencia permite el uso, distribución y reproducción del artículo en cualquier medio, siempre y cuando se otorgue el crédito correspondiente al autor del artículo y al medio en que se publica, en este caso, Medwave.