



Revista Peruana de Medicina Experimental y
Salud Pública

ISSN: 1726-4642

revmedex@ins.gob.pe

Instituto Nacional de Salud
Perú

Bolaños-Díaz, Rafael; Mezones-Holguín, Edward; Gutiérrez-Aguado, Alfonso; Málaga, Germán
LOS ESTUDIOS DE SÍNTESIS COMO BASE PARA LAS EVALUACIONES ECONÓMICAS:
NECESIDAD DE LA VALORACIÓN DE SU CALIDAD
Revista Peruana de Medicina Experimental y Salud Pública, vol. 28, núm. 3, julio-septiembre, 2011,
pp. 526-534
Instituto Nacional de Salud
Lima, Perú

Disponible en: <http://www.redalyc.org/articulo.oa?id=36321259019>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica
Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

LOS ESTUDIOS DE SÍNTESIS COMO BASE PARA LAS EVALUACIONES ECONÓMICAS: NECESIDAD DE LA VALORACIÓN DE SU CALIDAD

Rafael Bolaños-Díaz^{1,2,a}, Edward Mezones-Holguín^{1,3,b},
Alfonso Gutiérrez-Aguado^{1,4,c}, Germán Málaga^{5,d}

RESUMEN

Los estudios de síntesis (EDS): revisiones sistemáticas y metanálisis, son base para llevar a cabo las evaluaciones económicas en salud (EES). Los EDS, al permitir obtener parámetros para las probabilidades de estimación y de efectividad a partir de la combinación de resultados de estudios primarios y, al incluir en su metodología procesos de selección, evaluación, sistematización y síntesis, son considerados como el primer nivel de jerarquía en la evidencia científica. No obstante, pueden estar sometidos a sesgos y fallas metodológicas que afectan su validez. El presente artículo expone, inicialmente, la importancia de la aleatorización en la jerarquización de los diseños de investigación; luego se revisa los principales factores que afectan la validez de los EDS, incidiendo en el sesgo de publicación, la heterogeneidad y la inclusión de estudios primarios con objetivos principales distintos al del EDS; asimismo, se presenta a los estudios individuales como una alternativa válida para el desarrollo de una EES. Se concluye que uno de los aspectos claves en una EES es la selección adecuada de los tipos de estudio, sean estos primarios o secundarios.

Palabras clave: Revisiones; Metanálisis como asunto; Práctica clínica basada en la evidencia; Evaluación de la investigación en salud; Economía de la salud (fuente DeCS BIREME).

SYNTHESIS STUDIES AS THE BASIS FOR ECONOMIC EVALUATIONS IN HEALTH: THE NEED FOR THEIR QUALITY APPRAISAL

ABSTRACT

Synthesis studies (SS): systematic review and meta-analysis are the basis for developing Health Economic Evaluations (HEE). SS allow us to obtain parameters for estimating probabilities and effectiveness from the combination of the results of primary studies, and, as they include in their methodology the selection, evaluation, systematization and synthesis processes, they are considered the first level of hierarchy in scientific evidence. Nevertheless, they can be prone to bias and methodological failures that can affect the validity of their results. This article initially presents the relevance of the randomization in the hierarchic classification of research designs, then it reviews the main factors affecting the validity of the SS, emphasising the publication bias, the heterogeneity and the inclusion of primary studies with main objective differing from the one of the SS. Moreover, it presents individual studies like a valid alternative for the development of a SS. The conclusion is that one of the key aspects in a SS is the correct evaluation of the study types and the objective evaluation of their quality, being these primary or secondary.

Key words: Reviews, Meta-analysis as topic; Evidence-based practice; Health Research research evaluation, Health economics (source: MeSH NLM).

1. Unidad de Análisis y Generación de Evidencias en Salud Pública, Centro Nacional de Salud Pública, Instituto Nacional de Salud. Lima, Perú.

2. Escuela de Postgrado, Universidad Nacional Mayor de San Marcos. Lima, Perú.

3. Escuela de Medicina, Universidad Peruana de Ciencias Aplicadas. Lima, Perú.

4. Facultad de Ciencias Médicas, Universidad César Vallejo, Lima - Norte. Lima, Perú.

5. Centro De Excelencia CRONICAS, Universidad Peruana Cayetano Heredia. Lima, Perú.

^a Médico especialista en Medicina Interna, Magíster en Epidemiología Clínica, Magíster en Economía de la Salud; ^b Médico Cirujano, Magíster en Epidemiología Clínica; ^c Médico Especialista en Gestión en Salud; ^d Médico especialista en Medicina Interna, Doctor en Medicina

INTRODUCCIÓN

A pesar de las múltiples barreras que existen para traducir la evidencia científica desde la academia hacia los gestores y decisores ⁽¹⁾; las evaluaciones económicas son una herramienta relevante para la toma de decisiones en Salud Pública, por constituir un análisis que incluye tanto un componente de efectividad basado en la mejor evidencia disponible, así como, un análisis de costos contextualizados, para la implementación de diversas intervenciones sanitarias. En este marco los estudios de síntesis (EDS): las revisiones sistemáticas y el subsecuente metanálisis, cumplen una función importante, al ser un consolidado "ponderado" de estudios individuales con calidad evaluada que permiten obtener parámetros de estimación adecuados, incluyendo las probabilidades de ocurrencia o transición, así como información para el análisis de efectividad y/o utilidad. Es importante remarcar, que una revisión sistemática puede o no incluir metanálisis, el cual constituye la metodología estadística utilizada con el fin de obtener una medida cuantitativa de resumen de los estudios incluidos para una determinada variable de respuesta ^(2,3).

En los últimos años los EDS - también denominados estudios secundarios - han recibido un mayor interés tanto entre los investigadores, como en médicos y profesionales de la salud en general, debido a que ofrecen la ventaja de mostrar la línea de tendencia que siguen los resultados de los estudios primarios al respecto de un misma pregunta de interés. Por otro lado, debido a que utilizan métodos preespecificados para identificar, evaluar y sintetizar toda la evidencia disponible relacionada con una pregunta clínica, son ubicados en la cima de la jerarquía de la evidencia, situación que por lo general no es cuestionada, lo que motiva que los resultados dispuestos y mostrados por un EDS, sea prácticamente asumida como libre de dudas ^(4,5). Sin embargo, como cualquier tipo de estudio puede estar sometido a sesgos y por lo tanto constituir evidencia carente de validez. En ese sentido, resulta imperativo evaluar su calidad, más aun cuando puede ser base para el análisis de efectividad en una evaluación económica ^(1,2,3,6).

En el presente artículo, se presenta una breve introducción sobre la importancia de la aleatorización en la evaluación de relaciones causales y en la jerarquía de los estudios clínicos, posteriormente se expone una sucinta aproximación a los factores que pueden afectar la validez de un EDS para concluir en la importancia de la selección del estudio más adecuado para llevar a cabo una evaluación económica en salud.

LA ALEATORIZACIÓN Y LA JERARQUÍA DE LOS DISEÑOS DE INVESTIGACIÓN

Si bien un EDS puede incluir diversos tipos de investigaciones primarias, cuando se evalúa la efectividad de una intervención tanto terapéutica como preventiva, el estudio primario de elección a ser considerado para formar parte de una revisión sistemática es el ensayo clínico aleatorizado controlado (ECA) ^(2,7,8). La principal diferencia entre un ECA y un metanálisis radica en la aleatorización, proceso que confiere el carácter de experimento al estudio individual. En los diseños experimentales es el investigador quien sistematiza y asigna aleatoriamente la intervención a estudiar, siendo la experimentación uno de los métodos científicos que permite establecer relaciones de causalidad, las cuales constituyen la cualidad de mayor impacto cuando se jerarquiza la evidencia.

De igual modo, la aleatorización es la mejor herramienta para evitar el sesgo de selección, pues busca eliminar las diferencias entre los grupos de estudio que potencialmente pueden influenciar en los resultados observados, es decir, permite controlar en alguna medida el efecto de cualquier covariable oculta e indetectable que podría generar confusión entre los grupos de comparación ^(6,9-11). Es por ello, que el adolecer de falta de aleatorización sitúa a los estudios que carecen de este procedimiento en un segundo nivel de jerarquía en la evidencia clínica, justamente porque es mayor la posibilidad de no controlar cualquier variable oculta que pueda impactar en los resultados; a este nivel se le ha denominado clásicamente como "observacional analítico" (estudios de cohortes, casos y controles, transversales analíticos y transversales comparativos), donde se suma el hecho de que no existe una intervención sino una exposición. Si bien clásicamente suele considerarse que el análisis puramente observacional no permite establecer de modo directo relaciones de causalidad, sino asociaciones y correlaciones estadísticas, existen métodos epidemiológicos avanzados tanto en diseño y análisis que buscan mejorar esa aproximación ⁽¹²⁻¹⁵⁾.

Finalmente, se establece un tercer nivel de jerarquía compuesto por los estudios que al no tener el modelo de estudio suficiente robustez para responder a una pregunta de investigación analítica se le ha denominado "nivel descriptivo". Aquí se sitúan los estudios de prevalencia no analíticos, así como, las series y reportes de casos. Este nivel de jerarquía no soporta el rigor de los análisis matemáticos justamente porque el diseño del estudio no lo permite ^(7,14).

Por otro lado, existen diseños de estudio muy particulares que se ubican en una zona intermedia o "zona gris"

1° Nivel	Ensayo Clínico Aleatorizado Controlado	Intervención con asignación aleatoria
	Estudios cuasi-experimentales	Intervención sin asignación aleatoria
2° Nivel	Estudios Observacionales Analíticos	Sin intervención, con Exposición
3° Nivel	Estudios Descriptivos	Sin intervención, sin Exposición

Figura 1. Niveles Generales de Jerarquía en los Estudios Primarios.

entre el primer y segundo nivel, ya que cumplen con ciertos criterios de uno u otro. Un ejemplo de ello son los denominados estudios “cuasi-experimentales”, en los cuales no hay aleatorización, pero el investigador sistematiza y asigna la intervención a realizar. En este grupo, se encuentran las investigaciones con grupos paralelos cuya asignación no fue aleatoria, así como, los diseños antes-después donde los sujetos sometidos a una intervención forman parte también del grupo control; entre otros ^(12,13) (Figura 1).

Recientemente, algunos autores han cuestionado la supremacía jerárquica de los ECA al ser confrontados por los estudios observacionales, sustentado en las siguientes razones: 1) En algunas oportunidades, el principio de equiponderación (“*equipoise*”) de los ECA – que se basa en la incertidumbre que existe en torno a la diferencia de efecto entre las intervenciones – en la práctica no se aplica debido a que las intervenciones experimentales suelen ofrecer de antemano un mayor beneficio que el control, más aún cuando el grupo control recibirá el placebo; 2) En otras ocasiones, la aleatorización no es posible de llevarse a cabo por razones éticas y 3) ciertas veces la aleatorización puede representar un fenómeno de “auto-frustración” (“*self-defeating*”), sobre todo en estudios donde es necesario que el sujeto intervenido conozca el tratamiento para obtener la suficiente efectividad. En todas estas circunstancias, los estudios observacionales –fundamentalmente los estudios de cohortes – con diseño óptimo podrían ser una opción válida para suplir a los ECA ^(16,17).

Toda la jerarquización de la evidencia se hace considerando el potencial de sesgo de los modelos de estudio, es decir, el primer nivel de jerarquía estaría menos expuesto a sesgo considerando los criterios de buena calidad de estas investigaciones ⁽¹⁷⁾. Al respecto, a nivel de estudios primarios, sólo aquellos

ECA que reúnen criterios de calidad mínimos gozarán de esta menor exposición al sesgo. En 1995, Jadad *et al* desarrollan una primera aproximación de los criterios de mayor impacto en la validez interna de los ECA; tales como aleatorización, cegamiento, tasa de abandonos, análisis por intención a tratar, entre otros; este hecho introdujo la necesidad de contar con escalas efectivas para la valoración de calidad de estos estudios, las cuales han sido mejoradas progresivamente agregándose nuevas exigencias para liberarlos de sesgos ^(3,18).

Sin embargo, debido a que los EDS no cuentan con una aleatorización propia, puede surgir interrogantes como: *¿Son modelos experimentales? o ¿Por qué son considerados como primer nivel de jerarquía?* Se considera que el metanálisis tradicional constituye un modelo matemático muy elaborado que permite minimizar el sesgo y, al mismo tiempo, ponderar la validez de los estudios individuales que lo componen. Igualmente, concierne recordar que así como se sintetizan los resultados de ECA también puede sintetizarse los resultados de estudios observacionales; manteniendo su nivel de jerarquía. Al mismo tiempo, es importante exponer que la combinación de resultados de ECA (diseños experimentales) con estudios prospectivos de cohortes (diseños observacionales) es un modelo, al cual se recurre siempre que exista compatibilidad en los objetivos del EDS y siempre que cada estudio tenga la calidad suficiente para ingresarlo al análisis combinado. Si bien esta tendencia epidemiológica es reciente, sugiere que la combinación de ambos diseños permite ganar información clave al respecto de los objetivos de estudio ^(2,19,20).

PRINCIPALES FACTORES QUE AFECTAN LA VALIDEZ DE UNA REVISIÓN SISTEMÁTICA Y METANÁLISIS

Ya se mencionó que los EDS son considerados como el nivel más alto de jerarquía, sin embargo, en algunas oportunidades nos enfrentamos a revisiones sistemáticas constituidas por estudios clínicos con alto riesgo de sesgo, con búsqueda incompleta o desactualizada, reporte limitado o nulo de la concordancia entre evaluadores, análisis insuficiente o falta de reporte de heterogeneidad inapropiados, entre otros ^(21,22). En realidad son varios los factores que pueden afectar la validez de un EDS; sin embargo, nos centraremos los tres más importantes en el caso específico de los metanálisis: sesgo de publicación, heterogeneidad y uso de investigaciones primarias cuyo objetivo principal difiere del objetivo del EDS.

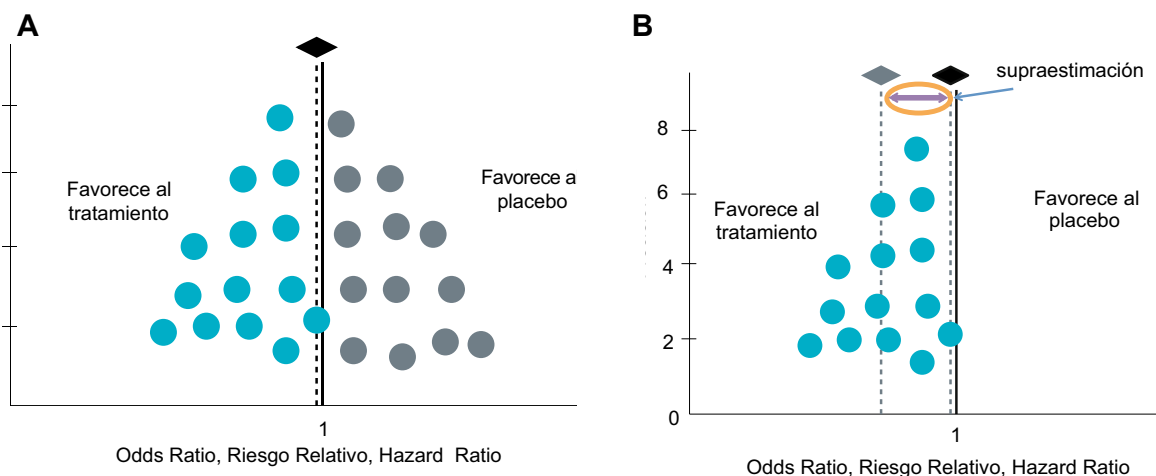


Figura 2. A. Caso hipotético con estudios publicados y no publicados, donde se observa que el Intervalo de confianza al 95% de la estimación del metanálisis incluye al 1. **B.** Caso hipotético que incluye solamente estudios publicados. Se observa que la estimación del metanálisis no incluye al 1, favoreciendo al tratamiento.

SESGO DE PUBLICACIÓN

El sesgo de publicación suele representarse gráficamente en el denominado diagrama de embudo (*"funnel plot"*) que confronta en un plano cartesiano la precisión con el resultado final de cada uno de los estudios, generando una dispersión de puntos cuyos contornos limitan la forma de un embudo, esto es debido a que los estudios con mayores tamaños de muestra presentan estimaciones más exactas que se concentran en un área de la dispersión de puntos, y los estudios con tamaños de muestra más pequeño tienen resultados con mayor variabilidad (mayor error estándar) por lo que éstos puntos se encuentran más dispersos^(23, 24) (Figura 2A y 2B).

El sesgo de publicación, se origina muchas veces por la arbitrariedad de las revistas biomédicas para decidir predominantemente la publicación de artículos que muestran diferencias significativas y positivas. Prefiriéndose estudios con muestras de gran magnitud, relegando a los estudios sobre pequeño número de pacientes, aunque éstos tengan un tamaño suficiente para responder su objetivo primario; situación que se ve acentuada, debido a que la metodología matemática inherente al Metanálisis analiza los ECA de acuerdo al error estándar de los resultados, otorgando mayor ponderación al estudio con estimación más precisa, vale decir aquellos con mayor tamaño de muestra^(2, 23, 24).

Hay varios métodos matemáticos que permiten detectar el nivel de este sesgo en los Metanálisis pero no son absolutos. Uno de ellos es la técnica de "Estudios bajo

la mesa" (*"File-Drawer Number"*), que tiene cierta similitud al Número Necesario a Tratar de los ECA; ya que evalúa el número de estudios con resultados opuestos suficientes para neutralizar o invertir la dirección del resultado global; si este número es muy alto será menor la probabilidad de cambiar el resultado brindándole mayor robustez estadística al metanálisis. Otro procedimiento es el denominado "Corte y relleno" (*"Trim and Fill"*) que busca reconstruir el diagrama de embudo con la simetría suficiente para obviar el sesgo; sin embargo, se requiere un gran número de estudios, una situación infrecuente en un EDS⁽²³⁾.

Aunque es importante completar la búsqueda sistemática con estudios no publicados y resúmenes de congresos, hay que tener en cuenta que, algunos expertos sugieren que en muchas ocasiones tales estudios no alcanzan una calidad metodológica suficiente para ser incluidos en la Revisión Sistemática. Debido a ello, consideran que es más conveniente invertir tiempo y dinero en actualizar la evidencia con las publicaciones más recientes que en buscar información no publicada. Bajo esta perspectiva convendría concentrarse más en incluir estudios mínimamente válidos, independientemente del tamaño de muestra; lo cual justamente por su gran variabilidad, tiene el potencial de robustecer estadísticamente los resultados finales del metanálisis. Sin embargo, sobre un mismo tema de estudio, varios EDS que han incluido estudios de reducido tamaño muestral y bajo poder estadístico coinciden suficientemente con los resultados de aquellos que únicamente han incluido estudios de alto poder estadístico, aún en presencia de sesgo de publicación. De aquí

se deriva que el punto crítico no debe focalizarse en la existencia o no de sesgo de publicación sino en el impacto final que éste tendría ⁽²⁵⁾. No obstante, se debe tener en cuenta y estar alertas con relación a potenciales inconductas científicas y/o conflictos de intereses, al dejar de publicar resultados no favorables, sobre todo cuando se trata de medicamentos ⁽²⁶⁾. Es importante enfatizar que en todo el proceso, la barrera del idioma no debería de representar inconveniente alguno.

HETEROGENEIDAD

La *heterogeneidad* es el otro problema inherente a los metanálisis, y se entiende como la variabilidad no explicada únicamente por el error de muestreo. Ello implica que cada estudio esté midiendo un efecto diferente de los otros, donde la varianza aparece a raíz de los diseños o poblaciones diferentes de estudio, adicionalmente a la varianza propia del muestreo. Ignorar la heterogeneidad en estas investigaciones implicarían infraestimar la varianza y el intervalo de confianza final, lo que aumenta la probabilidad de significancia estadística del resultado combinado sin que lo sea realmente. Se utilizan diferentes pruebas estadísticas para detectar el nivel de heterogeneidad en un metanálisis, siendo la metarregresión una de las pruebas más conocidas. Cuando los resultados provenientes de un metanálisis mantienen un nivel de sesgo de publicación y de heterogeneidad en rangos mínimos no significativos, se fortalecen sus conclusiones siempre que se haya respetado la metodología íntegramente ^(22, 27, 28).

USO DE ESTUDIOS PRIMARIOS CON OBJETIVO PRINCIPAL DISTINTO AL DEL METANÁLISIS

El metanálisis adquiere mayor contundencia cuando su objetivo principal coincide con el objetivo principal de las investigaciones individuales que lo componen. En ciertas oportunidades puede observarse que algunos de los estudios primarios que lo integran tienen como objetivo secundario lo que se ha definido como objetivo principal en el EDS; con lo que el modelo perderá finalmente validez interna en cierto grado.

Por ejemplo, el EDS llevado a cabo por Bolland *et al.* tuvo como objetivo principal evaluar el efecto de los suplementos de calcio sobre el riesgo de infarto agudo de miocardio y otros eventos cardiovasculares; sin embargo, los estudios que lo componían, en su mayoría este era un objetivo secundario ⁽²⁹⁾. Recientemente, Lewis *et al.* publicaron un ECA, con alto poder estadístico y doble cegamiento, donde el objetivo primario fue precisamente el desenlace cardiovascular ⁽³⁰⁾. Considerando lo reportado por ambas publicaciones, aparece una discrepancia en los resultados, con la disyuntiva de cuál di-

seño tendría mayor validez. Tras lo expuesto; el ECA al ser un estudio de alta calidad metodológica con menor exposición a la heterogeneidad— previa evaluación de su calidad - supera en contundencia estadística al metanálisis en mención, el cual a pesar de agrupar varios ECA, muestra la limitación de que los estudios primarios que lo conforman no tienen como objetivo principal el desenlace cardiovascular ⁽²⁹⁾.

CONCLUSIONES

Por lo general, las EES se sustentan en los resultados finales de uno o varios metanálisis, aprovechando que este diseño combina en un único resultado final los resultados de varios estudios sobre una misma pregunta de investigación. Sin embargo, en ocasiones aparecen discrepancias entre los resultados de un EDS y de un ECA de alto poder estadístico. Es por ello, muy importante conocer las características metodológicas que le brindan validez a ambos diseños, con lo que todo dependerá del análisis crítico que se realice.

Si el EDS cumple con los criterios de calidad indispensables será de elección frente a un ECA; sin embargo, si está debilitado por alguna de las razones que hemos señalado en los párrafos previos pues será preferible establecer la efectividad o eficacia en base al único estudio individual que se dispone. Cuando existen varios ECA suficientemente válidos para responder una pregunta de investigación, pero cuyos diseños no están adecuadamente estandarizados como para combinarse en un metanálisis, podemos obtener información de efectividad o eficacia tanto de uno como otro estudio ^(31,32).

En el caso de los EDS, se sugiere el uso de guías de evaluación crítica de la calidad, tales como AMSTAR o SURE (recomendada por la Evidence Informed Policy Network - EVIPNET de la Organización Mundial de la Salud, que incorpora también aspectos abordados por AMSTAR); entre otras ^(33, 34, 35). Es importante marcar la diferencia entre las escalas mencionadas que evalúan la calidad y el riesgo de sesgo de las Revisiones Sistemáticas y metanálisis, frente a las guías usadas para su reporte, como es el caso de PRISMA ⁽³⁶⁾.

Cuando no disponemos de información sobre eficacia provenientes de estudios experimentales, lo reportado por estudios observacionales - a pesar de las limitaciones de este tipo de investigaciones en cuanto a menor control del sesgo y su falta de aleatorización - que respondan adecuadamente a la pregunta de investigación planteada en el objetivo primario del estudio de síntesis pueden ser una alternativa pertinente.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica del Perú (CONCYTEC) por el apoyo brindado a EMH para su entrenamiento con la International Epidemiological Association (IEA) en tópicos relacionados a lo expresado en el presente artículo.

Contribuciones de autoría

RB redactó el primer borrador del artículo y realizó la revisión final. EM redactó la versión presentada al comité editorial, elaboró las figuras del artículo, levantó las observaciones del comité editorial y redactó la versión final del artículo. AG realizó contribuciones al texto inicial, desarrolló la primera versión del resumen y revisó la versión final. GM realizó contribuciones al texto inicial, levantó las observaciones del comité editorial, realizó contribuciones al texto final y revisó la versión final.

Fuente de Financiamiento:

Autofinanciado.

Conflictos de Interés

Los autores declaran no tener conflictos de interés en la publicación de este artículo.

REFERENCIAS BIBLIOGRÁFICAS

1. Sinclair DG. Evidence over barriers: important but not enough. *Healthc Pap.* 2011; 11(2):57-60.
2. Athanasiou T, Darzi A. Evidence Synthesis in Healthcare: A Practical Handbook for Clinicians. 1st Ed. Londres (Inglaterra): Springer 2011
3. Guyatt G, Meade M, Rennie D, Cook D. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. Second ed. New York: McGraw-Hill Professional Publishing; 2008.
4. Caldwell DM, Ades AE, Higgins JP: Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005; 331(7521):897-900.
5. Málaga G, Sánchez-Mejía A. Medicina basada en la evidencia: Aportes a la práctica médica actual y dificultades para su implementación. *Rev Med Hered.* 2009; 20(2):191-7.
6. Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions 5.0.1 [September 2008] Chichester, UK: John Wiley & Sons, Ltd; 2008.
7. Martín-Zurro A. Investigación en Atención Primaria. In: Martín-Zurro A, Cano-Pérez JF, eds. Atención primaria: Conceptos, organización y práctica clínica. Madrid: Elsevier; 2003.
8. Wang D, Bakhai A. Clinical Trials: A Practical Guide to Design, Analysis, and Reporting. Chicago: Redemedia; 2004.
9. Oxford Centre for Evidence-based Medicine. Levels of Evidence [March 2009]. London (UK): Centre for Evidence Based Evidence; 2009.
10. Greenberg R, Daniels S, Flanders D, Boring J, Eley JW. Medical Epidemiology. 3th ed. New York: McGraw-Hill Companies, Inc; 2002.
12. Roberts C, Torgerson D. Randomisation methods in controlled trials. *BMJ* 1998;317(7168):1301.
13. Schmoor C, Caputo A, Schumacher M. Evidence from nonrandomized studies: a case study on the estimation of causal effects. *Am J Epidemiol.* 2008 May 1;167(9):1120-9.
14. Schmoor C, Gall C, Stampf S, Graf E. Correction of confounding bias in non-randomized studies by appropriate weighting. *Biom J.* 2011;53(2):369-87.
15. Rothman K, Greenland S, Lash T. Modern Epidemiology (3rd Ed.). Philadelphia, PA: Wolters Kluwer Health / Lippincott Williams & Wilkins, 2008.
16. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health.* 2006; 60(7):578-86.
17. Marko N, Weil R. The role of observational investigations in comparative effectiveness research. *Value in Health* 2010; 13(8): 989-997.
18. Concato J. Observational versus experimental studies: what's the evidence for a hierarchy? *NeuroRx* 2004; 1: 341-347.
19. Jadad A, Moher D, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials* 1995; 16(1): 62-73.
20. Shrier I, Boivin JF, Steele RJ, Platt RW, Furlan A, Kakuuma R, Shrier I. et al. Should meta-analysis of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol* 2007; 166(10): 1203-9.
21. Kaizar E. Meta-analyses are observational studies: How lack of randomization impacts analysis. *Am J Gastroenterol* 2005; 100(6): 1233-6
22. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996; 15(24):2733-49.
23. Li T, Puhon MA, Vedula SS, Singh S, Dickersin K, Ahn-mmmwg AH. Network meta-analysis - highly attractive but more methodological research is needed. *BMC Med.* 2011;9(1):79.
24. Rothstein H, Sutton A, Borenstein M, eds. Publication bias in meta-analysis (1st Ed.). West Sussex, England. John Wiley & Sons, 2005
25. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detect by a simple, graphical test. *BMJ* 1997; 315(7109): 629-34.
26. Van Driel M, De Sutter, De Maeseneer J, Christiaens T. Searching for unpublished trials in Cochrane reviews may not be worth the effort. *J Clin Epidemiol* 2009; 62(8): 838-44.
27. Godlee F, Loder E. Missing clinical trial data: setting the record straight. *BMJ.* 2010;341:c5641.
28. Athanasiou, T; Darzi, A. Evidence Synthesis in Healthcare: A Practical Handbook for Clinicians. 1st Ed. Londres (Inglaterra), Springer 2011.
29. Cooper H, Hedges LV, eds. The handbook of research synthesis. Russell Sage Foundation, 1994.

30. Bolland M, Avenell A, Baron J, Grey A, MacLennan G, Gamble G, et al. Effect of calcium supplements on risk of myocardial infarction and cardiovascular events: meta-analysis. *BMJ* 2010; 341:c3691.
31. Lewis J, Calver J, Zhu K, Flicker L, Prince R. Calcium supplementation and the risks of atherosclerotic vascular disease in older women: Results of a 5-year RCT and a 4.5-year follow-up. *JBMR* 2011; 26(1): 35-41.
32. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between metaanalyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997; 21(337): 536-42.
33. DerSimonian R, Levine R. Resolving discrepancies between a meta-analysis and a subsequent large controlled trial. *JAMA* 1999; 282(7): 664-670.
34. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007; 7:10.
35. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol*. 2009; 62(10):1013-20.
36. Pantoja T. "Evaluando la calidad de las revisiones sistemáticas: SURE checklist". En: Taller de Elaboración de Policy Brief y preparación de diálogos deliberativos. Santiago de Chile, 7 -10 de Marzo del 2011. Washington DC: Evidence-Informed Policy Network (EVIPNET); 2011.
37. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med*. 2009;6(7):e1000100.

Correspondencia: Edward Mezones-Holguín
 Dirección: Av Cápac Yupanqui 1400, Lima 11.
 Teléfono: (511) 617-6200 anexo 2143
 Correo electrónico: emezones@gmail.com

Consulte las ediciones anteriores de la
 Revista Peruana de Medicina Experimental y Salud Pública en
www.scielosp.org

