

Medicina Basada en Evidencia: ¿podemos confiar en los resultados de los estudios clínicos aleatorizados bien diseñados?

A. Maturana • C. Benaglio

ABSTRACT

Evidence-based medicine: Can we trust the results of well-designed randomized trials?

Evidence based medicine assists in clinical decision-making by integrating critically appraised information with patient's values and preferences within an existing clinical context. A fundamental concept in this paradigm is the hierarchy of information. The randomized clinical trial is recognized as one of the designs that is less prone to bias and therefore of higher methodological quality. Clinical guidelines are one of the principal tools that evidence based medicine uses to transfer scientific information to clinical practice and many of their recommendations are based on these type of studies. In this review we present some of the limitations that the results can have, in even well designed and executed randomized clinical trials. We also discuss why valid results in these types of studies could not necessarily be extrapolated to the general population. Although the randomized clinical trial continues to be one of the best methodological designs, we suggest that the reader be careful when interpreting its results.

(**Key words:** Evidence-Based Medicine, clinical trial, clinical guideness, methodology).

Rev Chil Pediatr 2014; 85 (5): 533-538

RESUMEN

La Medicina Basada en Evidencia es una propuesta que asiste en la toma de decisiones clínicas integrando la información críticamente analizada con los valores y preferencias del paciente en el contexto clínico existente. Un concepto fundamental en este paradigma es la jerarquización de la información. El estudio clínico aleatorizado es reconocido como uno de los diseños metodológicos con menor probabilidad de sesgo y por ende de la más alta calidad metodológica. En este tipo de estudios se basan muchas de las recomendaciones de las guías clínicas, que son uno de los principales instrumentos que utiliza la medicina basada en evidencia para transferir la información a la práctica clínica. En esta revisión se exponen algunas de las limitaciones que pueden tener los resultados de estudios clínicos aleatorizados incluso cuando han sido bien diseñados y ejecutados.

Recibido el 15 de agosto de 2014, última versión aceptada el 1 de septiembre de 2014.

Andrés Maturana (✉)

Oficina de Desarrollo Educacional. Facultad de Medicina, Clínica Alemana - Universidad del Desarrollo.
Servicio de Neonatología, Clínica Alemana.
E-mail: amaturana@alemana.cl

Carla Benaglio

Oficina de Desarrollo Educacional. Facultad de Medicina, Clínica Alemana - Universidad del Desarrollo.

Se discute también el porqué resultados validos pueden no necesariamente ser extrapolables a la población general en este tipo de estudios. Si bien el estudio clínico aleatorizado sigue siendo uno de los mejores diseños metodológicos, se sugiere que el usuario de la información sea cuidadoso al interpretar sus resultados.

(Palabras clave: Medicina Basada en Evidencia, ensayo clínico, guías clínicas, metodología).

Rev Chil Pediatr 2014; 85 (5): 533-538

Introducción

La Medicina Basada en Evidencia (MBE) nace en la década de los noventa como una propuesta o paradigma que pretende acercar, de una manera ordenada, la información científica a la práctica clínica¹. Es frecuente que se piense que la MBE solamente comprende la evaluación crítica de la literatura científica, no obstante, es importante entender que esta herramienta asiste en la inevitable toma de decisiones clínicas que se requieren para practicar medicina. Si bien la información críticamente analizada y jerarquizada es uno de los pilares que propone la MBE, definitivamente no es el único². También hay que analizar el contexto clínico en el cual se enmarca esta decisión, además de considerar e integrar las preferencias y valores de aquella persona o grupo de personas que serán afectados por ella (figura 1)³. Este paradigma no pretende competir con la experiencia del clínico, que resulta fundamental para tomar la mejor decisión en base a los elementos antes mencionados⁴. En este artículo se pretende mostrar las fortalezas y limitaciones de un estudio clínico aleatorizado (ECA) en el contexto del paradigma de la MBE.



Figura 1.

Propuesta de la MBE para abordar problemas de salud

La propuesta que hace la MBE para abordar los problemas de salud a nivel poblacional o individual consiste en identificar los problemas clínicos, encontrar, jerarquizar y evaluar la información disponible para finalmente contextualizarla a la condición y valores del paciente o población³. Las guías clínicas son uno de los instrumentos que utiliza esta propuesta para lograr la transferencia de la información a la práctica clínica⁵. Una guía clínica es una recomendación desarrollada sistemáticamente para asistir al tratante y al paciente en la toma de decisiones en circunstancias clínicas específicas⁶. Estas son desarrolladas habitualmente por un grupo de expertos que analizan la información disponible y deciden sobre recomendaciones integrando el contexto en el cual se van a aplicar. Si bien algunas de sus recomendaciones se basan en información de alta calidad metodológica, otras lo hacen sólo en base al consenso de aquellos que las desarrollan^{5,6}.

Jerarquización de la información

Basado en lo anterior queda implícito que uno de los principios fundamentales en MBE es la jerarquía en la información científica. De esta manera, la "evidencia" o información puede asumir diferentes formatos que van desde las vivencias o experiencias clínicas no sistematizadas de un individuo o grupo clínico, hasta los estudios aleatorizados y las revisiones sistemáticas. La variable que define esa jerarquización es el sesgo o error sistemático y por ende, entre menos posibilidad de sesgo exista, mayor jerarquía va a tener la fuente de información o "evidencia". El sesgo es un error que

aleja o desvía los resultados sistemáticamente de la verdad subyacentes ya sea sobre o subestimando la magnitud de estos.

La experiencia clínica no sistematizada puede estar subjetivamente influenciada por aquellos resultados que más han impactado y por ende muy expuesta a sesgo. Los diferentes estudios clínicos, dependiendo del diseño metodológico, van a tener una mayor o menor exposición a sesgo por las limitaciones propias de su diseño. Si por ejemplo, comparamos datos recolectados prospectivamente en una cohorte de pacientes expuestos a una nueva terapia (grupo A en figura 2a), con un grupo control histórico no expuesto (grupo B en figura 2a), tendremos algunas limitaciones propias del diseño. Las cohortes comparadas van a diferir no sólo en la presencia o ausencia de la nueva terapia (grupos A y B en figura 2a). Debido a lo anterior, será imposible asegurar que cualquier diferencia encontrada en la evolución de estos dos grupos (X e Y en figura 2a), sea únicamente debida a la intervención. El solo hecho de que los grupos A y B estén siendo evaluados en momentos del tiempo diferentes (uno en el presente y otro histórico), podría explicar en parte las diferencias en evolución que podamos observar. Sin embargo, a veces esta es la única o la mejor información sistematizada disponible. En estos casos el paradigma de MBE propone leer críticamente esta informa-

ción para asegurar que dentro de las limitaciones del diseño no exista sesgo adicional debido a errores metodológicos. Con esa información evaluada críticamente en su validez, podemos enfrentar entonces la decisión clínica integrándola al contexto y valores del o los pacientes que estemos tratando.

Estudios clínicos aleatorizados

En la jerarquía de la información, uno de los diseños metodológicos que lidera, es el ECA. La razón fundamental es que en un ECA, bien diseñado y ejecutado, existe un gran control del error sistemático o sesgo dentro del estudio. Lo anterior permite lograr mayor validez de los resultados (validez interna) que en otros diseños. El ECA es por definición un experimento clínico en el cual logramos generar dos (o más) grupos con pronósticos similares mediante el proceso de aleatorización (figura 2b)⁷. Ambos grupos deberían al menos en teoría, lograr balancear la distribución de todas las variables pronosticas, es decir aquellas que impactan sobre la variable dependiente o resultado que se está estudiando (grupos A y B en figura 2b). Esto da como resultado una situación ideal a nivel metodológico, ya que logra controlar la única variable que los diferencia (variable independiente o intervención). De

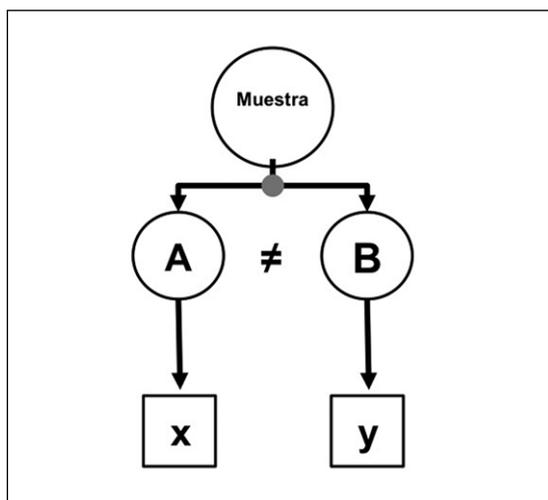


Figura 2a.

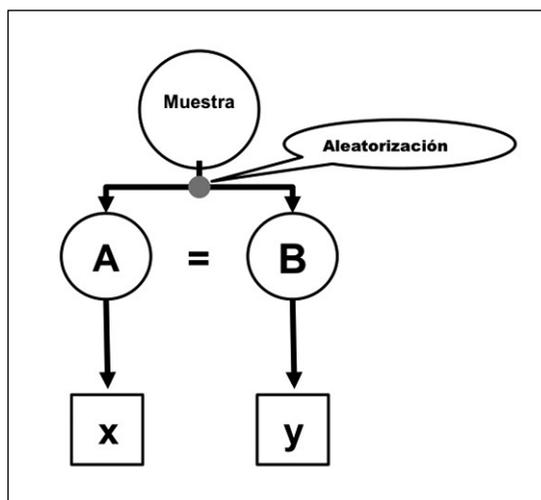


Figura 2b.

esta manera, si encontramos una diferencia en la variable estudiada entre los grupos (grupos X e Y en figura 2b) en la que podamos excluir razonablemente el azar, esta diferencia tendrá una sola explicación que será la intervención en estudio. Esto no lo vamos a lograr cuando comparamos dos grupos que no han sido generados aleatoriamente, y por definición van a ser diferentes en otras variables aparte de la intervención o exposición (figura 2a).

Limitaciones de un estudio clínico aleatorizado (ECA)

Al evaluar un ECA, es fundamental ver que estén presentes todos los elementos que garantizan la igualdad en variables pronosticas generadas entre los grupos comparados (grupos A y B en la figura 2b). Los detalles de esto se escapan de esta revisión pero están claramente definidos en la literatura⁷⁻¹⁰. El problema surge cuando tenemos una metodología bien diseñada y ejecutada pero existen otros elementos menos evidentes que comprometen esta credibilidad o validez. Vamos a ver a continuación algunos de esos elementos.

Impacto del azar

Cuando hablamos de que un resultado es “estadísticamente significativo” lo que estamos diciendo es que la probabilidad de que se deba al azar es menor al 5% ($p < 0,05$). Por definición consideramos esto como una probabilidad lo suficientemente baja como para desestimarla y asumir aquellos resultados como posibles o probables. Debemos estar conscientes que si bien esto hace poco probable el que sean sólo debido al azar, no lo excluye completamente y por lo tanto un resultado “estadísticamente significativo” no necesariamente constituye una verdad. Sin embargo, hay otras formas en las que el azar nos puede introducir dificultades en la interpretación de los resultados de un ECA. La aleatorización como explicamos anteriormente, nos permite generar (en teoría) dos grupos comparables en todas las variables pronosticas pero esto puede no siempre ser así.

Sólo por azar una variable pronóstica relevante puede quedar desbalanceada en forma importante y potencialmente introducir un sesgo en los resultados. Por ejemplo en el estudio de Morris y colaboradores donde se compara una estrategia de fototerapia profiláctica *versus* terapéutica en prematuros de extremo bajo peso, sucede un desbalance potencialmente importante en la variable sexo¹¹. Pese a que este es un estudio con una casuística de casi 1.000 pacientes por grupo y con una metodología impecable, se generó por azar una diferencia en el porcentaje de pacientes de sexo masculino entre ambos grupos, de 49 y 54% respectivamente. Esta diferencia de 5 puntos porcentuales no es menor ya que se trata de una de las variables pronosticas con mayor impacto en la prematuridad¹². Solamente por azar hay 5% más prematuros de sexo femenino (un factor protector), en el grupo que demostró mejor pronóstico en la variable principal de muerte o limitación neurológica, lo que introduce un potencial sesgo¹¹.

Fragilidad de los resultados

Otra de los elementos menos evidentes en los ECA es lo que podríamos llamar “fragilidad de los resultados”. Vamos a utilizar como ejemplo esta vez el estudio de Courtney y colaboradores, donde se compara la ventilación de alta frecuencia con la ventilación convencional en prematuros extremos¹³. En este estudio se demuestra una diferencia estadísticamente significativa a favor de la ventilación de alta frecuencia, hecho que los autores resaltan en sus conclusiones. Esta diferencia está dada por 131 neonatos de 234 (56%) con un resultado favorable en el grupo tratado, *versus* 117 de 250 (47%) en el grupo control. Sin embargo, si sólo uno de los 131 pacientes no hubiese tenido el resultado favorable (130 en lugar de 131), esta significancia estadística desaparece completamente. En este caso, la diferencia depende de un solo evento, lo que resulta extremadamente frágil y muy dependiente del azar. Nuevamente vemos un resultado favorable, en un estudio bien diseñado, que puede estar fuertemente amenazado por el azar.

Aplicación de los resultados

Si una terapia demuestra su efectividad en un ECA, habiendo descartado razonablemente la posibilidad de sesgo, la información debería ser extrapolable a una población similar a la incluida en el estudio². Esto parte de un supuesto, y es que la población que se incluyó en el estudio es una muestra representativa del universo de los pacientes que cumplen los criterios de ingreso para ese estudio. Sin embargo, por definición, los pacientes que participan en un estudio, no son una muestra al azar de la población (figura 3). No sólo deben cumplir con los criterios de inclusión, sino que también deben aceptar el participar en el estudio, de otra manera no se cumpliría con los principios básicos y fundamentales de la ética de investigación en seres humanos. Más aun, está bastante demostrado que el sólo hecho de participar en un estudio puede modificar el pronóstico de los participantes^{14,15}. Entonces debemos aceptar que la aleatorización protege del sesgo dentro del estudio, pero no evita que aquellos pacientes dentro del estudio se comporten de manera diferente de aquellos fuera del mismo (figura 3). Podemos utilizar como ejemplo el estudio SUPPORT, que evaluó dos rangos de saturación de oxígeno en prematuros extremos midiendo retinopatía severa del prematuro o muerte antes del alta¹⁶. Rich y cols., en una publicación posterior, analizaron las variables pronósticas de aquellos neonatos que cumplían los criterios de ingreso en el estudio SUPPORT, comparando aquellos que

habían participado con los que no lo hicieron en el mismo período de tiempo¹⁷. Este análisis demostró que los factores pronósticos eran significativamente diferentes entre ambos grupos comparados. Esto demuestra que el grupo que participo en el estudio era significativamente diferente al universo de pacientes con los mismos criterios de ingreso. El problema de este potencial sesgo de “representatividad” es que no es posible predecir su impacto ni existe manera de contrarrestarlo metodológicamente^{14,15}. Esto determina una limitación en la aplicación de los resultados de un ECA a la población general (figura 3).

Conclusiones

Si bien existen en la literatura varias pautas de cómo leer críticamente un ECA, estas se centran fundamentalmente en aquellos elementos de la metodología que protegen la aleatorización^{7-10,18,19}. La mayoría de estas recomendaciones pretenden ayudar a detectar errores metodológicos en el diseño o ejecución de los estudios. En esta revisión se ha mostrado que incluso ECA con un diseño metodológico impecable pueden no estar libres de sesgo. Los elementos analizados si bien a veces resultan menos evidentes, pueden comprometer tanto la validez interna (credibilidad) como la externa (aplicabilidad) de los resultados.

Probablemente el ECA sigue siendo uno de los mejores diseños metodológicos cuantitativos en la actualidad, sin embargo, el usuario de la información requiere ser muy crítico en su lectura.

Potenciales conflictos de interés: Este trabajo cumple con los requisitos sobre consentimiento /asentimiento informado, comité de ética, financiamiento, estudios animales y sobre la ausencia de conflictos de intereses según corresponda.

Referencias

- 1.- *Evidence-Based Medicine Working Group:* Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992; 268 (17): 2420-5.
- 2.- *Haynes RB, Sackett DL, Gray JM, Cook DJ, Guyatt*

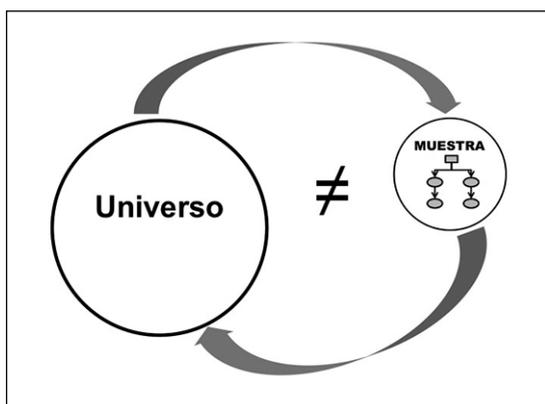


Figura 3.

- GH*: Transferring evidence from research into practice: 1. The role of clinical care research evidence in clinical decisions. *ACP J Club* 1996; 125 (3): A14-16.
- 3.- *Montori VM, Guyatt GH*: What is evidence-based medicine? *Endocrinol Metab Clin North Am* 2002; 31 (3): 521-6, vii.
 - 4.- *Haynes RB, Devereaux PJ, Guyatt GH*: Physicians' and patients' choices in evidence based practice. *BMJ* 2002; 324 (7350): 1350.
 - 5.- *Fried M, Quigley EMM, Hunt RH, et al*: Can global guidelines change health policy? *Nat Clin Pract Gastroenterol Hepatol* 2008; 5 (3): 120-1.
 - 6.- *Steinbrook R*: Guidance for guidelines. *N Engl J Med* 2007; 356 (4): 331-3.
 - 7.- *Peñaloza B, Candia R*: [Why is it essential to randomize a clinical trial on therapy?]. *Rev Med Chile* 2004; 132 (8): 1007-10.
 - 8.- *Pantoja T, Letelier LM, Neumann I*: [A critical appraisal of clinical reports in medical journals]. *Rev Med Chile* 2004; 132 (4): 513-5.
 - 9.- *Letelier LM, Manríquez JJ, Claro JC*: [Blinding for clinical trials. Does it matter?]. *Rev Med Chile* 2004; 132 (9): 1137-9.
 - 10.- *Capurro D, Gabrielli L, Letelier LM*: [Intention to treat and follow up are important in assessing validity of a randomized clinical trial]. *Rev Med Chile* 2004; 132 (12): 1557-60.
 - 11.- *Morris BH, Oh W, Tyson JE, et al*: Aggressive vs conservative phototherapy for infants with extremely low birth weight. *N Engl J Med* 2008; 359 (18): 1885-96.
 - 12.- *Tyson JE, Parikh NA, Langer J, Green C, Higgins RD, National Institute of Child Health and Human Development Neonatal Research Network*: Intensive care for extreme prematurity-moving beyond gestational age. *N Engl J Med* 2008; 358 (16): 1672-81.
 - 13.- *Courtney SE, Durand DJ, Asselin JM, et al*: High-frequency oscillatory ventilation versus conventional mechanical ventilation for very-low-birth-weight infants. *N Engl J Med* 2002; 347 (9): 643-52.
 - 14.- *McCambridge J, Witton J, Elbourne DR*: Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *J Clin Epidemiol* 2014; 67 (3): 267-77.
 - 15.- *McCambridge J, Kypri K, Elbourne D*: In randomization we trust? There are overlooked problems in experimenting with people in behavioral intervention trials. *J Clin Epidemiol* 2014; 67 (3): 247-53.
 - 16.- *SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network, Carlo WA, Finer NN, Walsh MC, et al*: Target ranges of oxygen saturation in extremely preterm infants. *N Engl J Med* 2010; 362 (21): 1959-69.
 - 17.- *Rich W, Finer NN, Gantz MG, et al*: Enrollment of extremely low birth weight infants in a clinical research study may not be representative. *Pediatrics* 2012; 129 (3): 480-4.
 - 18.- *Rada GG, Andrade AM*: [Should the results of this study be applied to my patient?]. *Rev Med Chile* 2006; 134 (1): 115-9.
 - 19.- *Rivera S, Larrondo FJ, Ortega JP*: [Interpreting the results of an article about therapy]. *Rev Med Chile* 2005; 133 (5): 593-6.