

### VALIDEZ INTERNA Y VALIDEZ EXTERNA DE UN ENSAYO CLÍNICO: EL EJEMPLO DEL ESTUDIO LIDO

Manel Pinteró, Pere Ventayol, Olga Delgado, Montse Vilanova

Servicios de Farmacia  
Hospital Universitario Son Dureta y Hospital Son Llàtzer. Palma de Mallorca

Para valorar si los resultados de un ensayo clínico son útiles en la práctica deben cumplir tres características: ser Válidos, Importantes y Aplicables (VIA)<sup>1-3</sup> (tabla 1).

#### A. Validez Interna

La validez interna plantea si los resultados obtenidos en un ensayo proporcionan una respuesta correcta al objetivo del estudio, con ausencia de sesgos en los procesos de diseño, ejecución y análisis. Un ensayo clínico «es válido» si permite asegurar que el efecto observado es debido a la diferente intervención que han recibido los grupos de estudio, y no a otros factores que puedan haber conducido a una falsa conclusión. En los estudios de superioridad hay tres aspectos básicos para valorarla:

1-Asignación aleatoria de los individuos. Es la característica metodológica clave. Significa que la decisión sobre qué tratamiento recibirá cada uno de los pacientes incluidos en los diferentes grupos de estudio se realiza al azar, de manera que se tiende a asegurar que se produce una distribución equilibrada de todas las variables pronósticas, tanto de las conocidas como de las desconocidas.

2-Seguimiento exhaustivo de todos los individuos. Dado que los grupos formados por asignación aleatoria tienden a ser comparables, todos los pacientes que los componen deben ser seguidos y tenidos en cuenta al finalizar el estudio. Si no se hace así, los individuos que finalizan el estudio pueden formar grupos que no sean similares, de forma que la comparación de las respuestas observadas no sea imparcial. No obstante, en la

Tabla 1.

Conceptos básicos para evaluar un ensayo clínico: VIA (validez, Importancia y aplicabilidad)<sup>1</sup>

- |    |  |
|----|--|
| V- | La validez interna depende principalmente de:<br>a. Asignación aleatoria de los individuos<br>b. Seguimiento exhaustivo de todos los individuos<br>c. Análisis por intención de tratar (ITT) |
| I- | La importancia clínica de los resultados depende de:<br>a. Tipo de variable utilizada<br>b. Magnitud de los resultados   |
| A- | La aplicabilidad de los resultados del ensayo a nuestros pacientes depende de:<br>a. El tratamiento<br>b. La población<br>c. La asistencia sanitaria   |

práctica siempre existe un número de pacientes que no pueden ser seguidos, por lo que es importante que en la publicación del ensayo clínico se proporcione información sobre los abandonos y las retiradas. Si sus características son diferentes de las de los individuos que finalizan el estudio, o si se han producido por motivos relacionados con las intervenciones del estudio, su exclusión del análisis podría invalidar las conclusiones.

3-Análisis por Intención de Tratar. La exclusión de determinados sujetos del análisis, por no haberse ajustado al protocolo o al seguimiento clínico definido en el estudio, puede estar relacionado con el pronóstico y ello podría introducir un sesgo de comparación. Para evitarlo, debe utilizarse la estrategia del análisis

por intención de tratar, según el cual en el análisis de resultados, los sujetos se incluyen en el grupo al que han sido asignados independientemente de si han recibido o no correctamente la intervención, seguimiento y evaluación<sup>10</sup>.

Otras características secundarias que orientan sobre la validez interna del ensayo clínico son:

**Enmascaramiento.** Las expectativas de éxito o de fracaso de la intervención, tanto por parte de los pacientes como de los investigadores, pueden influir en la evaluación de la respuesta. Para prevenirlo, se usan técnicas de enmascaramiento o ciego, que implican que los pacientes y/o los observadores desconozcan el tratamiento que recibe cada uno, de forma que no puedan interferir con los resultados. No siempre es posible aplicar estas técnicas, por ello como mínimo es recomendable que las personas que midan la respuesta desconozcan el tratamiento asignado, de forma que se reduzca el posible sesgo<sup>11</sup>.

**Comparabilidad inicial de los grupos.** Los grupos deben ser similares en todo, excepto en el tratamiento que reciben. Debe haberse evaluado la comparabilidad inicial de los grupos por los factores pronósticos, ya que la asignación aleatoria no garantiza que se haya producido una distribución equilibrada de todas las variables relevantes, especialmente cuando el número de individuos no es elevado. Si los grupos presentan algunas diferencias importantes por algún factor fuertemente asociado con el pronóstico, aunque no sea estadísticamente significativa, la validez del estudio puede verse comprometida. La comparabilidad de los grupos debe mantenerse a lo largo de todo el seguimiento.

Para llevar a cabo una revisión crítica de un ensayo clínico teniendo en cuenta todo lo anterior se han propuesto diferentes escalas y clasificaciones. Entre las más útiles y empleadas se encuentran la escala de A. Jadad<sup>12-14</sup>, y la lista-guía CONSORT<sup>15,16</sup>:

-La escala de Jadad es muy simple, fácil de aplicar y válida. Tiene 5 ítems que pueden sumar o restar puntos, y en una primera aproximación permite descartar los ensayos de baja calidad.

-La lista-guía CONSORT es muy completa y exhaustiva. Contiene un diagrama de flujo de pacientes y una lista con 22 apartados donde se revisan los puntos más relevantes de la validez del ensayo, sobre todo de validez interna.

## B. Validez externa y aplicabilidad de los resultados

La validez externa de un estudio ofrece información acerca de la generalización de sus conclusiones a

otros pacientes, que pueden ser similares a los del estudio o diferentes (en cuyo caso hablamos de extrapolación de los resultados).

Al valorar los resultados se analizará la variable estudiada y la magnitud e importancia de los resultados (diferenciando lo que es un valor estadísticamente significativo de lo que se considera como clínicamente relevante)<sup>17</sup>. Para determinar la aplicabilidad de los resultados, deben valorarse aspectos como la definición de la enfermedad, los criterios de selección de los sujetos, la intervención que se utiliza como referencia, la pauta y la dosis de las intervenciones que se comparan, y el entorno en el que se realizó el estudio.

El siguiente cuestionario es útil para valorar la validez externa y la utilidad práctica de los resultados del ensayo:

-¿Qué tipo de variable hemos medido? ¿Nos indica resultados clínicos? Si es una variable intermedia/subrogada, ¿está validada su asociación con resultados clínicos consistentes con el resultado final?

-¿Cuál es la relevancia clínica de los resultados?

-¿Es el fármaco con que se compara el estándar adecuado y el que se emplea en la práctica habitual de mi hospital?

-¿En caso de que el fármaco de comparación no sea el estándar de mi hospital, aporta el ensayo pivotal evidencias indirectas de utilidad?

-¿Las características basales de mis pacientes son similares a las del ensayo? ¿son muy restrictivos los criterios de inclusión y exclusión?

-¿El seguimiento clínico y asistencial de mis pacientes es similar al del ensayo?

-¿Se han tenido en cuenta todos los desenlaces clínicamente importantes?

-¿Los resultados pueden aplicarse a mi práctica?

## C. Ejemplo de evaluación de la validez del estudio LIDO<sup>18</sup>

### Evaluación de la validez interna del estudio LIDO

Escala de Jadad<sup>12-14</sup>. Como se observa en la tabla 2, el estudio LIDO cumple todos los requisitos con la puntuación máxima de 5.

Lista-Guía CONSORT<sup>15,16</sup>. El estudio LIDO cumple prácticamente todos los apartados: título y resumen, introducción, métodos, resultados y discusión. Solamente el apartado número 22, correspondiente a la interpretación de los resultados en función de la validez externa, podría considerarse dudoso.



# ÁREA TERAPÉUTICA

Tabla 2

ESCALA DE VALIDACIÓN DE ENSAYOS CLÍNICOS de A. JADAD <sup>12,14</sup> , aplicado al estudio LIDO <sup>10</sup>	PUNTUACIÓN
¿Se describe el estudio como aleatorizado? (*)	1
¿Se describe el estudio como doble ciego? (*)	1
¿Se describen los abandonos y exclusiones (*)	1
¿Es adecuado el método de aleatorización? (**)	1
¿Es adecuado el método de doble ciego? (**)	1
TOTAL	5
(*) SI= 1 / NO= 0. (**) SI= 1 / NO= -1 Rango de puntuación: 0-5. Estudio de baja calidad: Puntuación <3	

Globalmente y teniendo en cuenta los criterios anteriores, puede considerarse que el ensayo LIDO cumple con un alto nivel de validez interna.

## Evaluación de la validez externa y aplicabilidad del estudio LIDO

La siguiente encuesta (tabla 3) facilita el análisis de la validez externa del ensayo:

**Variable principal.** La variable principal empleada, «incremento del gasto cardíaco >30% y simultáneamente descenso presión capilar pulmonar de enclavamiento (PCWP)  $\geq 25\%$  a las 24 horas», es una variable intermedia y compleja. Se considera de relevancia clínica, aunque podría haber sido más idóneo emplear una variable como el «índice cardíaco».

**Importancia clínica de los resultados.** La diferencia absoluta de eficacia encontrada en la variable principal es del 13,2%; sin embargo, debe tenerse en cuenta que la eficacia del tratamiento inotrópico es moderada o baja (15% para dobutamina y 28,2% para levosimendán).

La mortalidad evaluada a los 31 días y a los 180 días presenta una reducción absoluta del orden del 10% pero es un objetivo secundario; se presenta en forma de curvas de supervivencia y se expresa en forma de Hazard Ratio (HR), con resultados de significación esta-

dística muy próximos al límite (ejemplo  $p=0,049$ ). Para unos mismos pacientes y una misma curva de supervivencia, se dan dos resultados de HR diferentes: uno a 31 días y otro a 180 días. Calculadas sobre valores de RAR, sólo es significativa la disminución de mortalidad a los 31 días, y con un IC 95% muy amplio.

Los tiempos de estancia en el hospital se expresan como «mediana del número de días con vida y fuera del hospital durante los primeros 180 días», y sugieren una menor estancia en el grupo levosimendán, pero los datos disponibles en la publicación de Cleland<sup>10</sup>, muestran que realmente no hay diferencias en la duración de la estancia hospitalaria entre ambos grupos.

Los resultados de subgrupos de pacientes en tratamiento con betabloqueantes son de tipo *post hoc*, por lo que pueden estar sometidos a sesgos importantes y no se deben sacar conclusiones (podrían plantear hipótesis de partida para nuevos ensayos).

**Comparador.** El comparador es dobutamina, que puede considerarse un estándar de tratamiento aceptable ya que es el tratamiento usual en nuestro medio, a pesar de los pocos estudios que avalen su eficacia. El problema está en la dosis y la duración del tratamiento utilizados en el ensayo LIDO. Dobutamina se ensaya a dosis de 5 mcg/Kg/min durante 24h y dosis máxima de 10 mcg/Kg/min, cuando en realidad en la práctica habitual se emplean dosis variables y con frecuencia durante periodos de más de 24 h. Como se observa en la figura 3 de la publicación del estudio<sup>10</sup>, al suspenderse la administración de dobutamina los parámetros hemodinámicos empeoran, hecho que también podría influir en los resultados de mortalidad encontrados.

**Tipo de pacientes.** La edad media (58-60 años) y las características clínicas y de gravedad de los pacientes son representativas del paciente tipo atendido en las unidades de cuidados intensivos. Sin embargo, no lo son de la mayor parte de aquellos con insuficiencia cardíaca descompensada que ingresan en los hospitales de nuestro ámbito. Estos suelen ser de edad más avanzada y de características clínicas diferentes.

## Conclusiones

El estudio LIDO es el principal ensayo clínico publicado que aporta evidencias sobre la eficacia y seguridad de levosimendán. Tiene un buen nivel de validez interna, pero presenta problemas de validez externa, sobre todo debido a la dosis y a la duración del tratamiento de dobutamina, con el que se compara, y al tipo de paciente estudiado.

**Tabla 3.**  
VALORACIÓN GLOBAL DE APLICABILIDAD  
DE LOS ENSAYOS CLÍNICOS: Aplicabilidad estudio LIDO

	Si/no
¿Es adecuada la variable utilizada?	Si
¿Son importantes clínicamente los resultados?	Si ↓
¿Es el tratamiento control adecuado en nuestro medio?	No
¿Nuestros pacientes son como los de la población estudiada?	Si/no
¿La práctica asistencial ensayada es factible?	Si

## Bibliografía

- Motaño Barrientos A, Rivas Aguiayo MD. Algunas consideraciones sobre la aplicabilidad de la MBE y la toma de decisiones en la práctica clínica diaria. *Atención Primaria en la red* ([www.fisterra.com](http://www.fisterra.com)).
- Jiménez Villa J. Lectura crítica de un ensayo clínico. Cuadernos de gestión 1997;3 (2): 76-89.
- Jiménez Villa J. Lectura crítica de la literatura científica (I) validez de estudio. *Formación Médica Continuada en Atención Primaria* 2000;7(4):209-220.
- Jiménez Villa J. Lectura crítica de la literatura científica (II). Evaluación de los resultados. *Formación Médica Continuada en Atención Primaria* 2000;7(5):283-296.
- Guyatt G, Pennie D, Hayward RSA. User's Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. AMA ed 2002. (<http://pubs.ama-assn.org/misc/usersguides.dtl>).
- Guyatt GH, Sackett DL, Cook DJ. User's guides to the medical literature. (II) How to use an article about therapy and prevention. A: Are the results of the study valid? *JAMA* 1993;270 (21):2598-2601.
- Guyatt GH, Sackett DL, Cook DJ. User's guides to the medical literature. (II) How to use an article about therapy and prevention. B: What were the results and will they help me in caring for my patients? *JAMA* 1994;271(1):59-63.
- García López FJ, Gutiérrez S, Galende I, Avendaño C. Evaluación de la calidad de los ensayos clínicos: justificación, utilidad e inconvenientes. *Med Clin (Barc)* 1999;112 (Supl 1):35-42.
- Gayoso P. Lectura Crítica de un artículo sobre tratamiento. Grupo MBE Galicia. [www.fisterra.com](http://www.fisterra.com)
- Absaira V. ¿Qué es el análisis por intención de tratar? *SEMERGEN* 2000;26:393-394.
- Rodríguez JL, et al. Doble ciego. El control de los sesgos en la realización de ensayos clínicos, contradicciones, insuficiencias e implicaciones. *Med Clin (Barc)* 2002 ;118: 192-195.
- Jadad AR, Moore A, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clinical Trials* 1996; 17 :1-12.
- Moher D, Jadad AR, Nichol G, et al. Assessing the Quality of Randomized Controlled Trials: An annotated Bibliography of Scales and Checklists. *Controlled Clinical Trials* 1995 ;16 :62-73.
- Jadad AR. Randomised Controlled Trials. BMJ Books 1998. 4ª reimpresión 2004. BMJ Publishing Group. ISBN 0-7279-1208-9.
- Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; 134: 683-694.
- Ioannidis JP, Evans SJ, Gøtzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141(10): 781-788.
- Fernández P, Pórtega S. Significancia estadística y relevancia clínica. *Cad Aten Primaria* 2001; 8: 191-195. ([www.fisterra.com](http://www.fisterra.com)).
- Follath F, Cleland JGF, Just H, et al. Efficacy and safety of intravenous levosimendan compared with dobutamine in severe low-output heart failure (the LIDO study): a randomised double-blind trial. *Lancet* 2002; 360: 196-202.
- Cleland JGF, et al. Intravenous levosimendan treatment is cost-effective compared with dobutamine in severe low-output heart failure: and analysis based on the international LIDO trial. *Eur J Heart Fail* 2003; 5 (1): 101-108.